
MGI

MESTRADO
Gestão de Informação

***Implementação de uma solução open
source de Business Intelligence com
metodologia Ágil***

Luís Pedro Lopes Batista

Proposta de Trabalho de Projeto apresentado como
requisito parcial para obtenção do grau de Mestre em
Gestão de Informação

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Implementação de uma solução open source de Business Intelligence com metodologia Ágil

por

Luís Pedro Lopes Batista

Proposta de Trabalho de Projeto apresentado como requisito parcial para a obtenção do grau de Mestre em Gestão de Informação, Especialização em Business Intelligence e Gestão de Informação na Saúde

Orientador: Professor Doutor Miguel de Castro Neto

DEDICATÓRIA

À Elsa e Sofia, para quem todas as palavras não chegariam.

AGRADECIMENTOS

Ao Professor Miguel Neto, pelo encorajamento e apoio constante para a realização deste trabalho.

Ao Dr. Paulo Vasconcelos (ACSS) e à Dr.^a Carla Santos (SPMS), pela ajuda e suporte incondicional na superação de obstáculos imprevistos.

Ao Dr. Luís Viana, pela motivação para ir mais além e não deixar de perguntar pelo porquê das coisas.

À Dr.^a Manuela Henriques, por muitas e grandes lições de vida.

RESUMO

A importância dos sistemas de data warehousing e business intelligence é cada vez mais pronunciada, no sentido de dotar as organizações com a capacidade de guardar, explorar e produzir informação de valor acrescido para os seus processos de tomada de decisão. Esta realidade é claramente aplicável aos sectores da administração pública portuguesa e, muito em particular, aos organismos com responsabilidades centrais no Ministério da Saúde. No caso dos Serviços Partilhados do Ministério da Saúde (SPMS), que tem como missão prover o SNS de sistemas centrais de business intelligence, o apelo dos seus clientes, para que possam contar com capacidades analíticas nos seus sistemas centrais, tem sido sentido de forma muito acentuada.

Todavia, é notório que, tanto os custos, como a complexidade, de grande parte destes projetos têm representado uma séria ameaça à sua adoção e sucesso. Por um lado, a administração pública tem recebido um forte encorajamento para integrar e adotar soluções de natureza open source (modelo de licenciamento gratuito), para os seus projetos de sistemas de informação. Por outro lado, temos vindo a assistir a uma vaga de aceitação generalizada de novas metodologias de desenvolvimento de projetos informáticos, nomeadamente no que diz respeito às metodologias Ágeis, que se assumem como mais flexíveis, menos formais e com maior grau de sucesso.

No sentido de averiguar da aplicabilidade do open source e das metodologias Ágeis aos sistemas de business intelligence, este trabalho documenta a implementação de um projeto organizacional para a SPMS, com recurso a ferramentas open source de licenciamento gratuito e através de uma metodologia de desenvolvimento de natureza Ágil.

PALAVRAS-CHAVE

data warehousing; business intelligence; open source; metodologias ágeis

ABSTRACT

The importance of data warehousing and business intelligence systems is ever more evident, as organizations strive to develop the ability to store, explore and produce value-added information that can enhance their decision-making processes. The reality of this situation is clearly applicable to the portuguese public administration sector and, in particular, the central agencies of the Health Ministry. In the case of the Health Ministry's Shared Services (HMSS) agency, tasked with the mission of providing centralized business intelligence systems throughout the health service, clients have manifested a very strong call to have their central systems fitted with analytic capabilities.

However, it is evident that both the cost and complexity of many of these projects has represented a serious threat to their adoption and success. On the one hand, the public administration sector has been strongly encouraged to adopt and integrate open source software (zero cost licensing models) into future development of IT projects. On the other hand, we have witnessed a wave of widespread acceptance of new development methodologies moving across IT projects, namely those that concern Agile methodologies, assumed as more flexible, less formal and more likely to succeed.

In order to ascertain the applicability of open source and Agile methodologies for business intelligence systems, this paper documents the implementation of a real-life project for HMSS, using open source tools (zero licence cost) and developed through Agile-natured methodologies.

KEYWORDS

data warehousing; business intelligence; open source; agile methodologies

ÍNDICE

1. Introdução	1
1.1. Discussão do contexto e identificação do problema	1
1.2. Objetivos do Projeto	4
1.3. Importância e relevância do estudo	4
2. Breve Enquadramento Teórico	6
2.1. Sistemas de Business Intelligence e Data Warehousing.....	6
2.1.1. Processos de ETL	7
2.1.2. Data Warehouse e Data Marts	10
2.2. Sistemas de Informação Open Source	14
2.2.1. O Conceito de Sistemas de Informação Open Source	14
2.2.2. A Abordagem Open Source aos Sistemas de Business Intelligence	16
2.2.3. Avaliação de Plataformas Open Source para Business Intelligence	16
2.2.4. Escolha de Plataformas Open Source de Business Intelligence	17
2.2.5. Plataformas Open Source de Business Intelligence (versão Comunidade). ..	18
2.2.6. Avaliação Global por Entidades Independentes.....	22
2.2.7. Comparação Final de Plataformas de BI Open Source	23
2.3. Metodologias Ágeis no Desenvolvimento de Sistemas de Informação	24
2.3.1. A Importância dos Sistemas de Informação	24
2.3.2. Metodologias e Processos de Desenvolvimento de SI	25
2.3.3. Evolução na utilização de metodologias e processos nos SI	25
2.3.4. As Metodologias Ágeis.....	29
3. Metodologia	31
3.1. Enquadramento Organizacional: a SPMS, EPE.....	31
3.1.1. Missão	31
3.1.2. Responsabilidade sobre os sistemas TIC do MS/SNS	32
3.1.3. A Lógica de Funcionamento da SPMS.....	33
3.1.4. Organização da SPMS	33
3.1.5. Linha de Serviços de Registos Centrais e Business Intelligence	34
3.2. Enquadramento Funcional: a Gestão de Projetos	34
3.2.1. Enquadramento Setorial da Gestão de SI pela SPMS	34
3.2.2. Os Projetos SI na SPMS	36

3.2.3. Gestão de Projetos SI na SPMS: Fundamento para o BI-GEST	37
3.3. Aspectos metodológicos do projeto BI-GEST	38
3.3.1. Âmbito.....	38
3.3.2. Processo	38
3.3.3. Metodologia.....	38
3.3.4. Infraestrutura.....	39
3.3.5. Software.....	39
4. Implementação	41
4.1. Iteração # 0 – Prova de Conceito	41
4.1.1. Desenho do Data Warehouse	42
4.1.2. Processos de ETL.....	44
4.1.3. Reporting.....	46
4.1.4. Plataforma de BI (Pentaho BI Server)	47
4.1.5. Resumo da Iteração # 0	48
4.2. Iteração # 1 – Aperfeiçoamento do Modelo de Dados.....	49
4.2.1. Desenho do Data Warehouse	49
4.2.2. Processos de ETL.....	50
4.2.3. Reporting.....	51
4.2.4. Plataforma de BI (Pentaho BI Server)	52
4.2.5. Análise de Resultados - Iteração # 1	54
4.3. Iteração # 2 – Implementação de Cubo Multidimensional	55
4.3.1. Pentaho Analysis Services (PAS)	55
4.3.2. Definição de schema do cubo multidimensional.....	55
4.3.3. Exploração visual do cubo com o Pentaho JPivot.....	57
4.3.4. Funcionalidades OLAP não utilizadas	59
4.3.5. Análise de Resultados - Iteração # 2	59
4.4. Iteração # 3 – Utilização de Ferramentas Melhoradas.....	60
4.4.1. Apresentação das Ferramentas CTools	60
4.4.2. Apresentação das Ferramentas Saiku.....	61
4.4.3. Instalação das Ferramentas	62
4.4.4. As Novas Ferramentas: Revisitando Reporting e Análise.....	63
4.4.5. Análise de Resultados - Iteração # 3	65
4.5. Iterações Futuras – Funcionalidades a Desenvolver.....	66
4.5.1. Camada de metadados	66

4.5.2. Camada de Dashboards	67
4.5.3. Slowly Changing Dimensions	68
5. Conclusões	69
5.1. Limitações	69
5.2. Objetivos Concretizados	70
6. Bibliografia	71

ÍNDICE DE FIGURAS

Figura 1 - Portais associados à Plataforma de Dados da Saúde (fonte: SPMS).....	2
Figura 2 - Framework para os sistemas de Business Intelligence (fonte: adaptado de Watson & Wixom, 2007).....	6
Figura 3 - Arquitetura Clássica de ETL (fonte: própria)	7
Figura 4 - Arquitetura funcional da plataforma Pentaho BI (fonte: i-Kno Knowledge Solutions)	19
Figura 5- Arquitetura funcional da plataforma comercial Jaspersoft BI (fonte: Jaspersoft)	20
Figura 6 - Arquitetura funcional da plataforma comercial Actuate BIRT (fonte: InfoDecisionnel)	21
Figura 7 - Arquitetura funcional da plataforma SpagoBI (fonte: SpagoBI).....	21
Figura 8 - "Quadrante Mágico" da Gartner para plataformas de BI - fev. 2013 (fonte: Gartner)	22
Figura 9 - Análise comparativa de plataformas open source de BI (fonte: Forrester Wave 3º trim. 2010).....	23
Figura 10 - As 5 Ondas de Inovação de Primozić (fonte: McNurlin, Spargue & Bui, 2009).....	24
Figura 11 - Sumário de resultados de projetos SI (fonte: Standish Group CHAOS Report 2010)	25
Figura 12 - Exemplo de processo em cascata (fonte: própria)	26
Figura 13 - Exemplificação do conceito de "Orientação Por Objetos" (fonte: própria)	28
Figura 14 – Comparação entre metodologias tradicionais e Ágeis (fonte: Standish Group CHAOS Report 2010)	30
Figura 15 - Mapa de Intervenção Estratégica para a Área de SI (fonte: ACSS 2010)	32
Figura 16 - Organograma da SPMS (fonte: SPMS).....	34
Figura 17 - Principais stakeholders no sector da saúde em Portugal (fonte: Deloitte 2011)	35
Figura 18 - Prestadores de cuidados de saúde em Portugal (fonte: Deloitte 2011)	35
Figura 19 - Amostra de projetos de SI da SPMS (fonte: SPMS 2013)	36
Figura 20 - As 10 Áreas de Conhecimento do Guia PMBOK (fonte: PMBOK)	37
Figura 21 - Formato de dados-fonte para utilização em prova de conceito (fonte: própria)	42
Figura 22 - Modelo lógico (star schema) para a primeira iteração (fonte: própria).....	43
Figura 23 - Modelo físico do data mart na primeira iteração (fonte: própria).....	43
Figura 24 - Processo de ETL no carregamento da tabela de Factos (fonte: própria).....	44
Figura 25 - Normalização do campo "dia", de colunas para linhas (fonte: própria)	45
Figura 26 - Atualização das datas de cada dia da semana (fonte: própria)	45
Figura 27 - Filtragem de datas sem registos de tempo (fonte: própria).....	46
Figura 28 - Esboço do relatório a produzir na iteração # 0 (fonte: própria)	46
Figura 29 - Elaboração do primeiro relatório na ferramenta Pentaho PRD (fonte: própria)	47
Figura 30 - Acesso ao relatório dentro do Pentaho BI Server (fonte: própria).....	48
Figura 31 - Diagrama do data mart para esta iteração (fonte: próprio)	49
Figura 32 - Processo de ETL melhorado, para carregamento de factos (fonte: própria)	51
Figura 33 - Configuração de conexão a base de dados no Administration Console (fonte: própria)	52
Figura 34 - Criação da fonte de dados "DW_iteration_1" para reporting (fonte: própria).....	53
Figura 35 - Construção de novos relatórios no WAQR (fonte: própria)	53
Figura 36 - Obtenção de relatório WAQR com listagem de colaboradores (fonte: própria)	54
Figura 37 - Janelas principais da ferramenta Pentaho Schema Workbench (fonte: própria)	56
Figura 38 - Representação do schema OLAP do projeto em XML (fonte: própria)	57
Figura 39 - Apresentação de análise padrão no JPivot (fonte: própria)	57
Figura 40 - Barra de ferramentas do Pentaho JPivot (fonte: própria).....	58
Figura 41 - Representação da funcionalidade de "drill through" (fonte: própria)	58

Figura 42 - Família de componentes na suite CTools (fonte: Webdetails)	61
Figura 43 - Principais funcionalidades da ferramenta Saiku (fonte: Saiku Analytics)	62
Figura 44 - Utilização do Marketplace para instalação de componentes (fonte: própria)	62
Figura 45 – Relatório com listagem de colaboradores através do Saiku Reporting (fonte: própria)	63
Figura 46 - Obtenção de matriz cruzada com o Saiku Analytics (fonte: própria).....	64
Figura 47 - Exploração gráfica dos dados no Saiku Analytics (fonte: própria)	64
Figura 48 - Exemplo de estatísticas básicas no Saiku Analytics (fonte: própria)	65
Figura 49 - Apresentação das 3 camadas do Pentaho Metadata Editor (fonte: Pentaho)	66
Figura 50 - Desenvolvimento de um dashboard no CDE (fonte: própria).....	67
Figura 51 - Desenvolvimento rápido de um dashboard de exemplo com CDE (fonte: própria)	68

ÍNDICE DE TABELAS

Tabela 1 - os 10 princípios do "Open Source Definition" (fonte: http://opensource.org/docs/osd)	15
Tabela 2 - Short-list de plataformas de BI open source (fonte: própria)	18
Tabela 3 - O Manifesto para Desenvolvimento Ágil de Software (fonte: agilemanifesto.org).....	29
Tabela 4 - Os 12 Princípios do Manifesto Ágil (fonte: agilemanifesto.org)	30
Tabela 5 - Discriminação dos principais componentes do Pentaho Analysis Services (fonte: própria)	55
Tabela 6 - Características do cubo multidimensional do projeto (fonte: própria)	56

LISTA DE SIGLAS E ABREVIATURAS

ACES	Agrupamento Complementar de Centros de Saúde
ACSS	Administração Central do Sistema de Saúde
AWS	Amazon Web Services
BI	Business Intelligence
CAP	Concursos Públicos de Aprovisionamento
CDA	Community Data Access
CDE	Community Dashboard Editor
CDF	Community Dashboard Framework
CE	Community Edition
DM	Data Mart
DW	Data WareHouse
ELT	Extract-Load-Transform
EPE	Entidade Pública Empresarial
ERD	Entity-Relationship Diagram
ETL	Extract-Transform-Load
ETLT	Extract-Transform-Load-Transform
GNU	GNU's Not Unix
GPL	General Public Licence
GUI	Graphical User Interface
IP	Instituto Público
JDBC	Java Database Connectivity
KPI	Key Performance Indicator
MDX	MultiDimensional eXpressions
MS	Ministério da Saúde
OLAP	Online Analytical Processing
PAS	Pentaho Analysis Services
PDI	Pentaho Data Integration
PMBOK	Project Management Body of Knowledge
PME	Pentaho Metadata Editor
PMI	Project Management Institute
PRD	Pentaho Report Designer
RDS	Relational Database Service

ROLAP	Relational OLAP
SCD	Slowly Changing Dimensions
SGBD	Sistema de Gestão de Base de Dados
SI	Sistemas de Informação
SNS	Serviço Nacional de Saúde
SPMS	Serviços Partilhados do Ministério da Saúde
SQL	Structured Query Language
TIC	Tecnologias de Informação e Comunicação
TR	Time Report
WAQR	Web Ad Hoc Query and Reporting
XML	eXtensible Markup Language

1. INTRODUÇÃO

O trabalho aqui apresentado resulta do desenvolvimento de um projeto de business intelligence para uma instituição do Ministério da Saúde, com o fim de prover essa mesma organização com uma camada analítica para os seus processos internos de gestão de projetos. Sendo o objetivo deste projeto uma solução informática destinada a uma utilização real, as opções conceptuais e práticas de trabalho adotadas durante a sua implementação – e aqui documentadas – não deixaram de ter igualmente em conta a submissão deste relatório para fins de apreciação académica.

1.1. DISCUSSÃO DO CONTEXTO E IDENTIFICAÇÃO DO PROBLEMA

Os sistemas de business intelligence têm vindo a ganhar cada vez maior importância no contexto do sector público da saúde em Portugal e, muito concretamente, no que concerne os sistemas de informação que suportam as operações diárias nas instituições do Serviço Nacional de Saúde (SNS). Muitos destes sistemas de informação, de cariz operacional e geridos localmente pelos diversos prestadores públicos de cuidados de saúde, são providos por uma entidade central (neste caso, a Administração Central do Sistema de Saúde, IP) através de um contrato-programa de prestação de serviços, estabelecido com a SPMS - Serviços Partilhados do Ministério da Saúde, EPE.

Na maioria de instituições do SNS, a “espinha dorsal” dos sistemas de informação de apoio à atividade de assistência clínica - o registo de atividade assistencial (SONHO/SINUS), sistemas de apoio ao médico e à enfermagem (SAM/SAPE), prescrição eletrónica de medicamentos (PEM), entre outros - é constituída por soluções que são da autoria e, portanto, da responsabilidade direta da SPMS. Para além das soluções informáticas destinadas a suportar e potenciar a assistência clínica, muitos dos sistemas utilizados nas funções de apoio e auxiliares - como serão os casos de sistemas de gestão de recursos humanos (RHV), de contabilidade (SIDC), etc. - também são providos pela SPMS.

Naturalmente, como forma de apoiar e facilitar a harmonização de utilização entre as diversas instituições do SNS, a SPMS conta com um pequeno (mas crescente) número de sistemas centrais, que possibilitam um armazenamento e controlo central sobre registos nacionais para um determinado conjunto de variáveis. São o caso, por exemplo, do registo nacional de utentes (RNU), registo nacional de profissionais (RNP), entre alguns outros. Para que a consulta e utilização destes recursos centrais fique

facilitada, encorajando a necessária e desejável interoperabilidade de sistemas transversais ao SNS, a SPMS promove o acesso a estes registos através da plataforma de dados da saúde (PDS). Note-se que este acesso pode ser efetuado por via humana (através de um de diversos portais, como é o caso do Portal do Utente), ou por via de sistema (recorrendo a uma sofisticada camada de web-services).

Figura 1 - Portais associados à Plataforma de Dados da Saúde (fonte: SPMS)



Para além da necessidade de assegurar o desenvolvimento, implementação e manutenção de todo este vasto conjunto de sistemas transacionais, disponibilizados localmente nas instituições do SNS, existem também necessidades centrais de exploração de informação; no fundo, trata-se de providenciar capacidades analíticas - ou seja, business intelligence - sobre o vasto leque de dados produzidos pelos sistemas locais de cada instituição. A SPMS é já responsável por um pequeno número de sistemas centrais de BI, essencialmente para as funções de controlo financeiro, contratualização de prestação de serviços nos hospitais EPE, e movimento assistencial dos cuidados de saúde primários (ACES/centros de saúde).

Todavia, a filosofia de desenvolvimento implícita a estes sistemas de business intelligence tem sido casuisticamente escolhida em cada um dos casos, pelo que a escolha de processo, abordagem metodológica e infraestrutura tecnológica é efetuada de forma ad hoc. Desta forma, cada um destes sistemas é desenvolvido de forma independente, pouco (ou nada) partilhando com sistemas análogos já existentes, ou ainda sob construção. Neste momento, o universo de soluções de business intelligence de âmbito central, é constituído por este conjunto de data marts isolados, concebidos sob prismas diferenciados e impossibilitados de uma utilização conjunta para efeitos de responder à necessidade de um data warehouse corporativo para o SNS.

Ao mesmo tempo que se debruça sobre este desafio, de harmonizar os sistemas centrais de business intelligence do SNS, colocando em marcha uma estratégia global que abarque o provimento de capacidade de BI para estes sistemas centrais, a SPMS sente necessidades próprias em termos de acompanhamento e monitorização da

gestão da sua atividade. Ou seja, e muito embora a atividade principal da SPMS seja dotar o SNS com sistemas operacionais e com plataformas de business intelligence, ela própria precisa de um sistema de BI sobre estes projetos - de certa forma, falamos de um meta-BI: um sistema de BI para analisar o desenvolvimento dos projetos da SPMS (que inclui construir sistemas de BI).

O problema sobre o qual se debruça este trabalho nasce precisamente desta necessidade dual: qual deve ser a estratégia a adotar quanto ao provimento de funcionalidades centralizadas de BI para o SNS, ao mesmo tempo que é preciso implementar capacidade de BI para a gestão interna dos projetos da SPMS, no âmbito de prestação de serviços de sistemas de informação para as instituições do SNS.

Dado que a segunda necessidade acima identificada - a de um sistema de BI para a gestão de projetos na SPMS - é claramente de teor mais pragmático do que a primeira - uma estratégia para os sistemas centrais de BI no SNS - parece-nos que seja possível utilizar a resolução desta segunda necessidade como forma de alavancar e clarificar a eventual resolução da primeira. Ou seja, a SPMS pode socorrer-se da implementação do seu sistema de BI interno, para testar e experimentar algumas das ideias que poderão revelar-se úteis na escolha da estratégia a adotar para os sistemas centrais de BI do SNS.

Muito especificamente, no que concerne sistemas de BI, há dois importantes fatores que a SPMS há muito deseja clarificar: o grau de usabilidade e adequabilidade de sistemas open source para soluções de business intelligence, e a possibilidade de utilização de metodologias mais flexíveis na implementação destes sistemas. O problema que o trabalho aqui apresentado aborda é, precisamente, o de responder a estas duas questões:

1. É possível implementar um sistema (não-trivial) de business intelligence, concebido para uma utilização organizacional, através de uma plataforma open source que não apresente custos de licenciamento?
2. É possível implementar um sistema (não-trivial) de business intelligence, concebido para uma utilização organizacional, através de uma metodologia de natureza Ágil (ou seja, mais flexível e menos formal), em que o desenvolvimento possa ser realizado de forma iterativa e incremental?

1.2. OBJETIVOS DO PROJETO

Os objetivos deste trabalho são, assim, a conceção e implementação de um sistema de business intelligence que auxilie a gestão de projetos na área de prestação de serviços de sistemas de informação na SPMS.

Este sistema deverá ser construído com recurso a uma plataforma de BI open source, em que o modelo de licenciamento seja totalmente grátis, para que a SPMS possa auferir experiência sobre a atual oferta deste tipo de plataformas e dos respetivos níveis de desempenho, funcionalidades e capacidade de execução.

Simultaneamente, este projeto de BI deverá ser desenvolvido com base em metodologias Ágeis, cumprindo os princípios refletidos na Manifesto Ágil, como forma de permitir à SPMS a avaliação dos efeitos práticos destas metodologias mais informais e flexíveis, em contexto de projetos de pequena-média dimensão.

1.3. IMPORTÂNCIA E RELEVÂNCIA DO ESTUDO

Não restam dúvidas da importância que os sistemas de business intelligence terão no futuro do SNS, como o atestam as palavras do Sr. Ministro da Saúde em conferência realizada a 6 de novembro de 2012 (Portal da Saúde, 2012):

“Se para gerirmos bem o SNS necessitamos de mais e melhor informação para a gestão, atempada, comparada, correta e auditável, é na informação clínica que reside o conhecimento organizacional do nosso cliente final, o cidadão.

Ao contrário do que acontece muitas vezes, ainda demasiadas vezes, o ideal é determos informação dos nossos cidadãos não já depois de doentes mas antes de ficarem doentes. Esse será o futuro dos sistemas de informação na Saúde.

Um SNS que não conhece bem os seus cidadãos utilizadores, que não trabalha para delinear e colorir uma fotografia dinâmica, tão completa e atualizada quanto possível dos mesmos, é um serviço que não dará a melhor resposta clínica e da forma mais eficiente - ambos imperativos morais de um serviço de cariz social, financiado com dinheiro público.”

Por outro lado, também não restam dúvidas de que as soluções open source estão recomendadas para adoção pela administração pública como forma de estímulo ao crescimento económico, “sempre que a maturidade e o custo sejam favoráveis” (Resolução do Conselho de Ministros n.º 12/2012) sendo de notar que a recomendação do Plano Global Estratégico de Racionalização e Redução de Custos nas TIC na Administração Pública é de que:

“... devem os organismos públicos, antes de adquirirem qualquer solução ou serviços tecnológicos, proceder sempre à quantificação do Total cost of ownership, face aos requisitos mínimos definidos e comunicados ao mercado e custos (diretos e indiretos) da solução. Nesta análise, será obrigatória a comparação de soluções baseadas em software sujeito a licenciamento e de software aberto.”

Aliás, deve ser referido que a proposta de Orçamento de Estado para 2014 preconiza, de forma categórica, a preferência por software open source (PL 387/2013 - Proposta de Lei para o Orçamento de Estado para 2014):

“Artigo 6.º

Utilização das dotações orçamentais para software informático

1 -As despesas com aquisição de licenças de software, previstas nas rubricas «Software informático» dos orçamentos dos serviços integrados e dos serviços e fundos autónomos, apenas podem ser executadas nos casos em que seja fundamentadamente demonstrada a inexistência de soluções alternativas em software livre ou que o custo total de utilização da solução em software livre seja superior à solução em software proprietário ou sujeito a licenciamento específico, incluindo nestes todos os eventuais custos de manutenção, adaptação, migração ou saída.”

Por fim, será curial lembrar que no que concerne o desenvolvimento de sistemas de informação, e muito em particular sistemas de business intelligence, as taxas de insucesso são ainda muito alarmantes, face ao valor do investimento global nestes projetos (Saran). Aliás, cremos que não se trata apenas do custo “afundado” no caso de insucesso, mas talvez de maior relevância, o custo de oportunidade por não se ter conseguido providir a organização com o desejado sistema de BI. Não havendo soluções “mágicas” (Brooks, 1987), veremos que há evidência que aponta na escolha de metodologia de desenvolvimento como sendo um dos aspetos fulcrais na obtenção de sucesso nestes projetos. Desse modo, a importância de ganhar experiência com novas abordagens metodológicas - mais flexíveis, menos formais e mais focadas na interação com o cliente - será também um aspeto determinante do trabalho aqui apresentado.

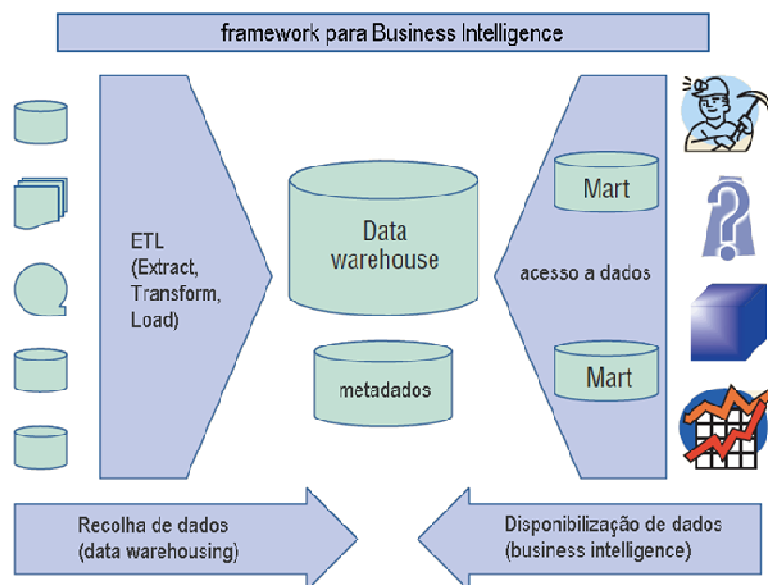
2. BREVE ENQUADRAMENTO TEÓRICO

Cremos ser indispensável abordar, sucintamente, os principais tópicos cobertos por este projeto, nomeadamente o business intelligence, os sistemas de informação de open source e as metodologias de desenvolvimento de filosofia Agile. Dado o objeto fundamentalmente prático deste trabalho, não se pretende uma exposição exaustiva do atual estado da arte nestas temáticas, mas sim, focar as principais ideias com relevância para este projeto.

2.1. SISTEMAS DE BUSINESS INTELLIGENCE E DATA WAREHOUSING

O conceito de Business Intelligence está definido de forma relativamente dispersa, podendo o seu significado abranger diversas interpretações, consoante as áreas de intervenção e estudo. Todavia, entre as interpretações de sentido mais lato, em que BI é descrito como os “processos que suportam aplicações de processamento analítico” (Watson & Wixom, 2007), às definições mais específicas, com BI a ser “sistemas que combinam a recolha e armazenamento de dados, com a gestão do conhecimento e ferramentas analíticas para apresentar informação interna, de elevada complexidade e representando vantagens competitivas aos responsáveis pelo planeamento e processo de tomadas de decisões nas organizações” (Negash, 2004), é possível perceber que BI representa uma framework de referência para todo este projeto, conforme se apresenta na Figura 2:

Figura 2 - Framework para os sistemas de Business Intelligence (fonte: adaptado de Watson & Wixom, 2007)



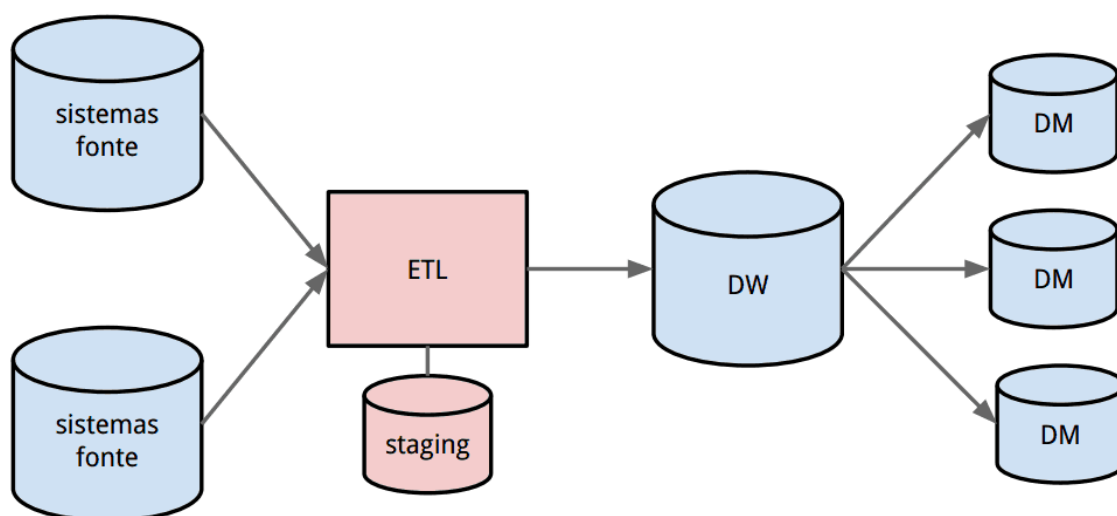
De seguida, destacaremos alguns aspetos importantes, no que concerne a conjugação destes componentes com a premissa base de testar e avaliar a adequabilidade de sistemas de BI open source.

2.1.1. Processos de ETL

Uma parte integrante dos projetos de data warehousing prende-se com a necessidade de obter os dados a partir dos sistemas fonte, transformá-los adequadamente para que sirvam os propósitos do DW previamente (e cuidadosamente) modelado e, por fim, carregar os dados já transformados para as estruturas do DW (tipicamente, as tabelas de dimensões e factos). Este processo é habitualmente designado de “ETL”, do inglês Extraction-Transform-Load (extração, transformação e carregamento).

Importa reforçar a importância deste processo, que pode representar uma parte muito significativa do esforço global que o projeto de DW/BI exigirá; Demarest aponta, por exemplo, para a possibilidade de até 80% do esforço total estar representado pelos processos de extração, tratamento e carregamento dos dados (Demarest, 1997). Para além destas tarefas, que são previsíveis e que podem ser planeadas, o mesmo autor refere, ainda, a hipótese do processo de ETL revelar problemas ‘escondidos’ nos dados a utilizar pelo projeto, ou a descoberta de que dados necessários não estão a ser capturados pelos sistemas transacionais. Por tudo isto, é crítico que o processo de ETL receba uma consideração ponderada e séria, quando do planeamento global do projeto.

Figura 3 - Arquitetura Clássica de ETL (fonte: própria)



Conforme referem Bouman & Dongen (2009, p. 224), as três principais tarefas no processo de ETL são:

1. **Extração.** A aquisição de dados, a partir de um ou mais sistemas fontes (que não obrigatoriamente sistemas de gestão de base de dados). Por exemplo, localizar e extrair todos os registos de vendas desde a última extração, no sistema operacional de suporte a vendas.
2. **Transformação.** A adequação que é operada sobre a estrutura ou forma dos dados, permitindo que sejam inseridos nas estruturas específicas do data warehouse. Por exemplo, a pesquisa e respetiva inserção de uma chave específica para cada linha de registo de vendas, consoante a zona regional de venda.
3. **Carregamento.** A operação de preenchimento das tabelas no respetivo data warehouse, com os dados que foram alvo de transformação.

Será de notar que existem algumas variantes referentes a este processo de ETL, nomeadamente no que concerne a ordem e frequência de cada uma das etapas anteriormente descritas. Nesse sentido, podemos falar de:

ETL (vulgo ETL Clássico): extração, seguido de tratamento e só no fim o carregamento dos dados. É o paradigma clássico de ETL, e tipicamente obriga a que os dados passem pelo motor / ferramenta de ETL, onde ocorre a transformação dos mesmos.

ELT: extração de dados, que são diretamente carregados para o sistema de destino, onde se realizam as transformações. Tipicamente, um sistema de gestão de base de dados executa essas transformações (que pode ser na fonte, no destino, ou mesmo em ambos os pontos, em determinados casos).

ETLT: extração dos dados, para fora dos sistemas fonte, seguido de transformações por ferramenta de ETL e carregamento para o sistema de destino, onde os dados são novamente alvo de transformações finais (tipicamente, as que são mais eficientes de realizar em contexto de base de dados).

▪ **Como escolher uma ferramenta de ETL?**

Dada a importância do processo de ETL nos projetos de business intelligence, há hoje em dia uma grande variedade de ferramentas específicas, dedicadas à extração, tratamento e carregamento de dados. A escolha por uma determinada ferramenta de ETL, num determinado projeto, deve obedecer a critérios objetivos, maximizando a eficácia e eficiência que a ferramenta adequada acrescentará ao projeto.

Apresentamos 9 aspetos que devem ser tidos em conta (Imhoff, Galemme, & Geiger, 2003, p. 286):

1. **Acesso aos dados.** Essencialmente, trata-se de perceber quais são as fontes de dados e em que tipos de sistemas de gestão de dados (base de dados, tipicamente) estão os dados-fonte. Pode a ferramenta conectar-se a estas fontes e que impacto terão estas conexões sobre o sistema operativo? Será também curial perceber como gere a ferramenta a transferência de dados na rede/domínio inerente ao DW.
2. **Volume de execução.** A preocupação com a eficiência da ferramenta, na extração e tratamento dos dados (principalmente, nos casos de volumes significativos), não deve ser descurada. Trata-se de saber se a ferramenta efetua boa utilização de cache / memória (especialmente em ações de lookup e chaves estrangeiras), se executa transformações de forma síncrona ou assíncrona, se pode beneficiar de múltiplos processadores, se suporta paralelização de processos., etc.
3. **Extensibilidade.** É importante perceber se a maior parte das transformações que serão necessárias num determinado projeto de DW/BI estão cobertas pela ferramenta; no caso de ser necessário adaptar, ou mesmo desenvolver transformações novas, a ferramenta permite-o? A ferramenta de ETL prevê a reutilização de componentes lógicos em transformações diferentes, ao longo de todo o projeto?
4. **Aquisição de dados em tempo real.** O projeto em consideração exige a captura e utilização de dados em tempo real (ou quase tempo real: *near real time*)? Se for este o caso, a ferramenta de ETL oferece estas funcionalidades?
5. **Meta dados.** Como referem Imhoff, Galemme, & Geiger (2003, p. 287), o desenho e implementação de um processo de ETL traduz-se na construção de ações, filtros e transformações dos dados que capturam uma grande variedade de regras de validação e de negócio (as denominadas business rules), valiosíssimas para as restantes partes do projeto de DW/BI. Será muito relevante ter em conta a capacidade da ferramenta de ETL para documentar, disseminar e tornar acessível estes meta dados a outras camadas do projeto (por exemplo, à camada de OLAP ou de análise ad hoc).
6. **Ambiente de desenvolvimento da ferramenta.** Qualquer ferramenta que venha a integrar o fluxo de desenvolvimento do projeto de DW/BI deverá ser avaliada em termos de integração face ao paradigma utilizado pela equipa e

pelos programadores de ETL, e face a outras ferramentas já conhecidas. O interface da ferramenta agrada à equipa de projeto? Permite a reutilização de transformações, componentes e outros artefactos do processo de ETL? Integra-se com as atuais práticas da equipa, ao nível de controlo de código fonte, versionamento e metodologias de desenvolvimento (nomeadamente, metodologias ágeis e releases contínuas)? É fácil de usar ao nível de debugging e testes?

7. **Ambiente de projeto.** Estas preocupações centram-se na integração da ferramenta sob consideração, com o atual ambiente em que está a ser desenvolvida a solução de DW/BI. Nomeadamente, devem ser avaliadas questões como o volume de dados que será alvo de ETL, a capacidade e caracterização da infraestrutura de rede afeta ao projeto, necessidades de integração com sistemas ou módulos externos, as exigências ao nível de infraestrutura por parte da própria ferramenta de ETL, entre outras.
8. **Prova de conceito.** Nos casos onde não existe experiência prévia com a ferramenta que está sob consideração, deve ser considerada a hipótese de realizar-se uma prova de conceito (eventualmente, coordenada e/ou executada pelo fornecedor da ferramenta), para que seja possível aquilatar das potencialidades da ferramenta num contexto semelhante ao previsto para o projeto real.
9. **Fornecedor / fabricante.** Como em qualquer aquisição de produtos e soluções informáticas, devem ser considerados alguns aspetos específicos no que concerne o fabricante (ou fornecedor) da ferramenta, nomeadamente em termos de oferta de treino e formação, suporte pós-venda, compromisso respeitante à melhoria e desenvolvimento futuro da ferramenta e, não menos importante, a estabilidade estrutural da empresa (ou organização) responsável pela ferramenta.

2.1.2. Data Warehouse e Data Marts

O conceito de data warehouse remonta aos final dos anos 80, onde era tido como sendo “o único repositório lógico de toda a informação de negócio utilizada para efeitos de reporting” (Devlin & Murphy, 1988). Todavia, terá sido Bill Inmom - conhecido como o “pai do data warehousing” - que popularizou o conceito de data warehouse, e que providenciou os critérios da definição hoje comumente aceites (Inmom, 2005). O data warehouse é um repositório de dados:

- i. **Orientado por assunto** - todas as entidades e transações relacionadas com um determinado assunto - por exemplo, vendas, recursos humanos ou hospitais - estão ligadas entre si.
- ii. **Integrado** - Inmom refere que este é, de facto, o aspeto mais importante de um data warehouse (Inmom, 2005); os dados que alimentarão o DW são oriundos de vários sistemas fonte, e por isso necessitam de ser limpos, filtrados e transformados, para que possam representar uma visão integrada do negócio. Por exemplo, convertendo diversas medidas diferenciadas (centímetros, polegadas, metros, quilómetros) para um referencial comum (em metros, por exemplo).
- iii. **Não-volátil** - os dados que são carregados para o data warehouse, por norma, não sofrem alterações ou atualizações. Ao contrário dos sistemas operacionais, apenas durante as operações de carregamento (periódicas e planeadas) é que os dados são alvo de algum tipo de edição.
- iv. **Variável no tempo** - ao contrário de um sistema transacional, em que os registos apenas estão associados ao momento de acesso (são válidos 'agora'), um data warehouse mantém registos que estão associados a todos os momentos onde ocorreu uma alteração nos dados (dados históricos). Ou seja, um data warehouse pretende manter 'fotografias' dos dados, de tal modo que seja possível aceder aos dados que eram válidos, em cada um desses momentos.

▪ **Vantagens de um DW**

Há vantagens óbvias em dotar a organização de um DW, sendo algumas delas as seguintes (Bouman & Dongen, 2009):

- 1. **Toda a informação está contida num só lugar** - torna desnecessário o esforço de tentar reunir informação espalhada por diversos sistemas, ficheiros ou mesmo mensagens de correio eletrónico; para além de ter a informação já reunida, está também já integrada, pelo que não é necessário realizar operações morosas e complexas, apenas com o objetivo de permitir 'cruzar' a informação.
- 2. **A informação está atualizada** - como o carregamento do DW é efetuado periodicamente, de acordo com um plano previamente delineado, a informação estará sempre devidamente atualizada.
- 3. **Acesso rápido** - a natureza técnica inerente ao DW possibilita, e privilegia, rapidez nas consultas que são executadas sobre o mesmo; certamente, com

maior velocidade de resposta de que as fontes alternativas (sistemas operacionais, ficheiros, email, etc.).

4. **Sem limites de volume** - os limites de volume num DW são, por norma, muito menos limitativos de que os meios convencionais (ficheiros, etc.). Em casos especiais - DW corporativos, big data, etc. - podem ser provisionados arquiteturas que possibilitam a utilização de um DW como se fosse, para o utilizador comum, 'infinito' em termos de volume de armazenamento.
5. **Disponibilidade de dados históricos** - se desenhado e implementado de forma adequada, o DW possibilita ao utilizador saber não só os dados do momento atual, mas também os dados de qualquer momento no tempo (desde que haja dados para esse intervalo temporal).
6. **Fácil de perceber** - dado que o DW é organizado de acordo com os assuntos pertinentes ao negócio (vide ponto 1 acima), e a informação está já integrada e acessível com base em descrições naturais (sem recurso a acrónimos ou a códigos internos), é significativamente mais fácil usar um DW de que recorrer a consultas sobre sistemas transacionais.
7. **Definições claras e consistentes** - o conceito subjacente ao DW, bem como o seu processo periódico de carregamento (ETL), fazem com que a informação que se pode extrair do DW esteja uniformizada e alinhada em termos de conceitos de negócio; a preparação dos dados, para que possa ser carregada para o DW, tem o efeito simultâneo de os normalizar, quanto a definições e conceitos de negócio.
8. **Dados estandardizados** - para além dos conceitos e normas subjacentes aos dados de que beneficia um DW (todos compreendem e revelam o mesmo entendimento de "lucro", presente no DW), também há uma desejável estandardização dos próprios dados (por exemplo, o referido conceito de "lucro" está medido em "euros", através de todo o DW - e não em euros numa tabela, milhares de euros noutra, etc.).

Será de notar que as primeiras 5 vantagens acima referidas dependem muito do rigor, competência e cuidado técnico associado ao desenvolvimento da respetiva solução de data warehouse. Estas 5 vantagens podem ser medidas, quanto ao seu grau de sucesso, apenas de um ponto de vista estritamente tecnológico. Já os últimos três pontos dependem muito mais do envolvimento e participação dos futuros utilizadores - mesmo no caso onde os primeiros 5 pontos foram tecnicamente assegurados, o DW

pode representar um projeto falhado, nos casos onde os utilizadores sentem que as últimas 3 vantagens não foram atingidas.

▪ **Data Warehouse vs Data Marts**

No que concerne os conceitos mais específicos de Data Warehouse e Data Mart, será oportuno referenciar as duas principais filosofias comumente aceites hoje em dia, conhecidas como a escola "Top-Down" de Inmon (Inmon, 2005) ou a escola "Bottom-Up" de Kimball (Kimball & Ross, 2002).

Ambas as filosofias reconhecem a importância do conceito de “repositório especial” para suportar as tarefas de exploração e reporting de uma determinada organização. A diferença reside na forma como cada uma destas duas escolas considera que este repositório especial deve ser estruturado. De acordo com Bouman & Dongen (2009, p. 115), há 3 pontos diferentes na abordagem ao data warehousing:

- i. ***Data Warehouse versus Data Marts com Dimensões Conformadas*** - a escola “Top-Down” de Inmon defende que o data warehouse deve ser único e central, abrangendo todos os assuntos que possam interessar à organização. Em contraste, a escola “Bottom-Up” de Kimball introduz o conceito de data marts - pequenos repositórios, dedicados em exclusivo a um só assunto (vendas, recursos humanos, etc.) - e da exploração destes data marts através de dimensões conformadas (características ou atributos comuns aos vários assuntos). Não iremos detalhar muito mais estes conceitos de dimensões, factos e os respetivos esquemas que lhes estão associados (em estrela, em cristal de neve, etc.), mas importa reter que é algo associado à filosofia Kimball de data warehousing.
- ii. ***Abordagem centralizada versus abordagem iterativa / descentralizada*** - no caso da escola de Kimball, o perímetro do data warehouse é composto por um conjunto integrado de data marts, ao contrário do entendimento da escola Inmon, que considera que apenas há lugar a um data warehouse central (sem prejuízo da utilização de data marts a juzante, para efeitos de eficiência na exploração dos dados). Note-se que no caso Inmon, os problemas inerentes à normalização, segurança e fiabilidade histórica dos dados são resolvidos num só lugar (o DW central) enquanto no caso Kimball, estes problemas precisam de ser resolvidos em cada data mart.
- iii. ***Modelo de Dados Normalizado versus Modelo de Dados Dimensional*** - mais do que as duas diferenças anteriormente identificadas, o principal ponto de discórdia entre as duas escolas prende-se com o modelo de dados

que cada escola preconiza para o DW. No caso da filosofia Inmon, defende-se a utilização de um modelo relacional, adequadamente normalizado (tipicamente, 3ª Forma Normal), para que a eficiência de armazenamento e referência cruzada entre as entidades do DW sejam maximizadas. No caso da filosofia Kimball, defende-se um modelo alternativo - designado de modelo dimensional - em que as tabelas de dimensões armazenam toda a informação, em formato desnormalizado, que interesse para efeitos de caracterização das métricas armazenadas nas tabelas de factos.

2.2. SISTEMAS DE INFORMAÇÃO OPEN SOURCE

2.2.1. O Conceito de Sistemas de Informação Open Source

A ideia de sistemas de informação de fonte aberta, ou seja, em que o acesso ao código fonte seja garantido, e com poucas ou nenhuma restrições à sua utilização e modificação, não é recente. De acordo com Lerner & Tirole, podem apontar-se três eras de evolução no conceito de open source:

1. A primeira era: do início dos anos 60 ao início dos anos 80

Trata-se de um tempo em o desenvolvimento de software ocorria em meios académicos, como era o caso da Universidade de Berkeley ou do Massachusetts Institute of Technology (MIT), ou em organizações comerciais em que os investigadores e programadores tinham um elevado grau de autonomia (por exemplo, Bell Labs e a unidade de investigação da Xerox em Palo Alto). Nesta fase, uma das principais preocupações era o desenvolvimento colaborativo de sistemas operativos que pudessem funcionar em diversas plataformas distintas - tendo no sistema Unix o seu exemplo mais conhecido. A cultura destes tempos encoraja a livre partilha de código fonte, de forma informal e ainda sem grande distinção na separação entre código fechado e código aberto.

2. A segunda era: do início dos anos 80 aos início dos anos 90

No início desta era começaram a surgir os primeiros problemas resultantes da ingenuidade com que se tinha abordado a livre partilha de código fonte de software importante, com a AT&T a querer assegurar propriedade intelectual sobre Unix. Uma das figuras desta era seria Richard Stallman, que em 1983 cria a Free Software Foundation, introduzindo um modelo formal de licenciamento de software open source, conhecido como General Public Licence (GPL) e motivando o desenvolvimento do sistema operativo GNU (versão Unix, livre da

interferência da AT&T). Essencialmente, os pioneiros do licenciamento open source GNU introduziram 4 garantias na utilização do software (GNU):

- Liberdade completa na forma como escolhe utilizar o software;
- Liberdade completa para distribuir gratuitamente o software;
- Liberdade para poder alterar o software, através de acesso integral ao código fonte;
- Liberdade completa para redistribuir gratuitamente versões melhoradas do software;

O licenciamento GPL introduzia a garantia adicional de que nenhum utilizador ou implementador de software open source podia revogar ou condicionar estes direitos, no caso da redistribuição de software de génese GPL (ou seja, que tivesse integrado alguma parte de código com licenciamento GPL).

3. A terceira era: do início dos anos 90 aos tempos atuais.

Com o surgimento e crescimento da internet, esta era conheceu uma aceleração notória na utilização de software open source. A escala de contribuições para projetos opens source aumentou significativamente, sendo o caso do Linux um exemplo bastante paradigmático do início desta era. Progressivamente, os modelos de licenciamento foram também sendo refinados, com a licença GPL a perder progressivamente a quase exclusividade de que gozava neste campo. Nos anos 90, em particular, organizações como a Debian (distribuidora de uma versão de Linux) contribuíram com novas formas de assegurar uma maior flexibilidade na utilização de software open source. Esta iniciativa acabaria por culminar na “Open Source Definition”, o manifesto atualmente tido como de maior expressão e aceitação, em termos de definir o que é “software open source”:

Tabela 1 - os 10 princípios do "Open Source Definition" (fonte: <http://opensource.org/docs/osd>)

1. Redistribuição livre	6. Nenhuma discriminação setorial / atividade
2. Acesso ao código fonte	7. Distribuição da licença
3. Trabalhos subsequentes	8. Licença não pode ser específica a produto
4. Integridade do código fonte do autor	9. Licença não pode restringir outros produtos
5. Nenhuma discriminação contra pessoas	10. Licença tem de ser tecnologicamente neutra

Hoje em dia, a utilização de software open source está muito disseminada, estando presente em muitos e diversos campos de interesse, incluindo a área de business intelligence e analytics, conforme demonstramos na seção seguinte.

2.2.2. A Abordagem Open Source aos Sistemas de Business Intelligence

É globalmente reconhecido por diversas entidades com responsabilidades de monitorização e estudo das tendências de mercado que a oferta de soluções de Business Intelligence, integradas no conceito de código aberto, tem melhorado muito substancialmente nos últimos anos:

"... past 5 to 7 years [...] OS BI products evolved tremendously, becoming increasingly viable options to their commercial counterparts. They can now be rightfully considered as valid alternatives"

(Faubert, 2012)

"These [OS] vendors don't just offer competitive options [...] they offer nearly fully functional broad BI suites that include reporting, OLAP, data visualization, dashboards, and data integration functionality"

(Evelson & Hammond, 2010)

Todavia, será importante referir que, derivado da complexidade das plataformas de Business Intelligence, da dificuldade inerente à integração dos diversos componentes aplicacionais e de aspetos concorrenciais, todas as atuais soluções que possam ser consideradas viáveis para projetos empresariais apresentam um modelo de negócio assente numa oferta dicotómica:

- **Versão Comercial:** uma plataforma open source mas de natureza comercial (com custos de utilização), em que tipicamente são oferecidos componentes exclusivos e/ou melhorados, níveis de integração e administração significativamente melhorados e de utilização facilitada, e com oferta integrada de suporte, formação e documentação;
- **Versão "Comunidade":** uma plataforma open source totalmente livre de custos de utilização, em que a integração entre o conjunto de componentes que constitui a suite de Business Intelligence pode variar substancialmente, e não há lugar a suporte técnico;

2.2.3. Avaliação de Plataformas Open Source para Business Intelligence

Uma das principais vantagens frequentemente referida na adoção de soluções open source, sejam ou não de Business Intelligence, é o baixo custo no investimento inicial (especialmente quando comparado com soluções proprietárias). Nem sempre a oferta de Business Intelligence em open source fica a custo-zero: há muitas escolhas ao nível de suporte, oferta melhorada, integração e ferramentas de administração e suporte. Ainda assim, os custos iniciais de aquisição são significativamente mais baixos e podem ser, nos casos onde se escolhe versões Comunidade, praticamente zero.

Deste modo, como avaliar e escolher entre as diversas ofertas de pacotes open source de Business Intelligence? Note-se que esta escolha traduzirá, obrigatoriamente, um compromisso importante em termos de alocação estratégica de recursos organizacionais, pois os menores níveis de suporte (especialmente nas versões Comunidade), exigem mais das equipas técnicas afetas aos projetos de Business Intelligence. Neste contexto, não é recomendável dispersar esforços e recursos organizacionais entre diversas plataformas distintas, pela que a importância da escolha inicial sai reforçada.

No âmbito deste documento, e em consonância com as necessidades de avaliação da SPMS, iremos considerar apenas uma short-list das plataformas de BI open source que estão disponíveis no mercado, tendo em conta os seguintes vetores estratégicos:

- **Oferta atual:** cada um dos candidatos a considerar deve apresentar qualidade mínima na sua oferta ao nível de arquitetura da solução, ambientes de desenvolvimento de soluções de BI, e apresentar níveis razoáveis de capacidade funcional e operacional nas suas plataformas;
- **Estratégia:** tentaremos ter em conta também o posicionamento estratégico de cada uma das plataformas e dos seus principais sponsors, em termos do seu compromisso para manter e continuar a desenvolver a plataforma de BI;
- **Presença no mercado:** por fim, será também tido em conta a atual posição de cada uma destas plataformas no mercado de Business Intelligence, ao nível de quota de mercado, relevância financeira e comercial, níveis de integração vertical e horizontal, entre outros.

2.2.4. Escolha de Plataformas Open Source de Business Intelligence

Com base nos vetores acima referidos, foram estabelecidos critérios de avaliação comparativa ao nível de cada plataforma open source de Business Intelligence considerada neste documento, nomeadamente:

- Abrangência de Componentes Incorporados na suite de BI: idealmente, cada uma das plataformas de Business Intelligence deve garantir a oferta dos componentes principais de ETL, reporting operacional, análises interativas (ad hoc), OLAP e dashboards;
- Deve permitir análises através do paradigma relacional (linguagem SQL) ou multidimensional (linguagem MDX), sendo ideal que disponha de ambos os

modelos (queries alternativas - como, por exemplo, XQuery, XPath, NoSQL, etc. - não foram valorizados);

- Integração da oferta num único ambiente de Business Intelligence, auto-suficiente e com boa capacidade operacional ao nível de processos e workflows de BI (desde administração e tarefas de back-office, à capacidade de agendamento de reporting e alertas);
- Capacidade de demonstrar que está a ser utilizado em cenários reais, em ambientes de produção e com feedback positivo e encorajador, por parte dos respetivos utilizadores e sponsors;

Desta forma, foi elaborada uma short-list de plataformas open source de Business Intelligence que aparentam estar alinhadas com estes critérios e exigências estratégicas, que são as seguintes:

Tabela 2 - Short-list de plataformas de BI open source (fonte: própria)

	Identificação Plataforma	Organização responsável	Oferta Comercial (Enterprise)	Oferta Comunidade (Community)
1	Pentaho BI	Pentaho, Inc.	Pentaho Business Analytics	Pentaho BI Suite CE
2	Jaspersoft BI	Jaspersoft Corp.	Jaspersoft BI Enterprise	Jaspersoft BI Community
3	BIRT	Actuate	Actuate BIRT	Eclipse BIRT
4	SpagoBI	SpagoBI (Ingegneria Informatica S.p.A.)	n/a	SpagoBI

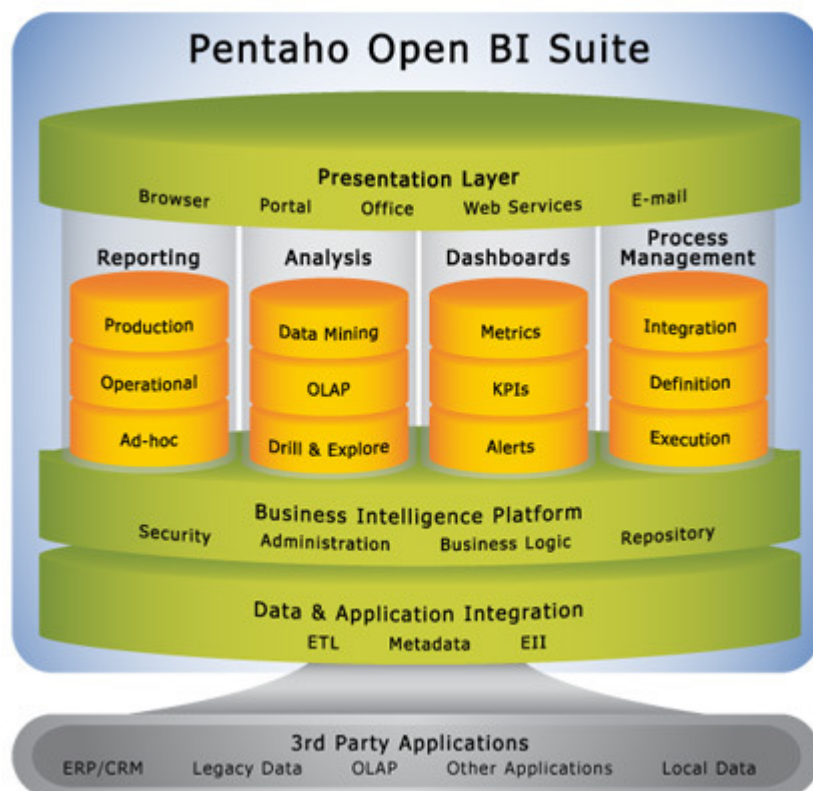
De seguida, avaliaremos individualmente as principais características de cada, focando essa análise apenas nos aspetos diferenciadores ou que, de algum modo, possam limitar a sua utilização para efeitos de projetos futuros da SPMS.

2.2.5. Plataformas Open Source de Business Intelligence (versão Comunidade)

▪ Pentaho BI Suite CE

Uma das mais completas ofertas de suite integrada de BI, com uma grande abrangência de componentes para as mais variadas fases do processo de Business Intelligence, conforme se apresenta na seguinte figura:

Figura 4 - Arquitetura funcional da plataforma Pentaho BI (fonte: i-Kno Knowledge Solutions)



A suite de BI Pentaho CE conta com ferramentas ou componentes distintos mas relativamente integrados para as principais funcionalidades, sendo que a operacionalidade básica em algumas destas camadas é bastante limitado. Em termos concretos, ao nível de análises interativas (ad hoc) e dashboards, a oferta inicial é muito fraca (havendo componentes muito melhorados para estes fins na oferta comercial). Todavia, existem um conjunto de ferramentas construídas pela comunidade (CTools, Saiku Analytics, etc.) que melhoram substancialmente as capacidades nativas da plataforma, deixando-a a um nível comparável com outras plataformas nestas funcionalidades específicas.

Em termos de reporting puro (orientado para relatórios pré-concebidos e respetiva distribuição por via de vários canais em diversos formatos: html, xml, xls, csv, etc. por via de web, email, etc.) a solução da Jaspersoft será melhor. Todavia, a solução Pentaho beneficia de melhor integração entre os seus componentes e a sua edição Comunitária está mais dotada para responder às necessidades transversais de sistemas BI - desde o ETL (ferramenta específica) a portal de dashboards.

▪ Jaspersoft BI Community

A suite de BI disponibilizada pela Jaspersoft continua a ser o marco de referência ao nível de reporting (por via das ferramentas próprias JasperReports e Jaspersoft

iReports Designer), tendo sido a escolha de muitas organizações em projetos de grande dimensão. Para os restantes componentes de BI, esta organização adota soluções externas open source, escolhendo entre as opções disponíveis, para depois as integrar como componentes próprias do seu pacote global:

Figura 5- Arquitetura funcional da plataforma comercial Jaspersoft BI (fonte: Jaspersoft)



O calcanhar de Aquiles desta solução prende-se com a sua oferta na versão Comunidade (sem custos de aquisição / licenciamento), em que há algumas limitações ao nível de componentes de dashboards e análises interativas - estes componentes não estão disponíveis nesta versão da suite.

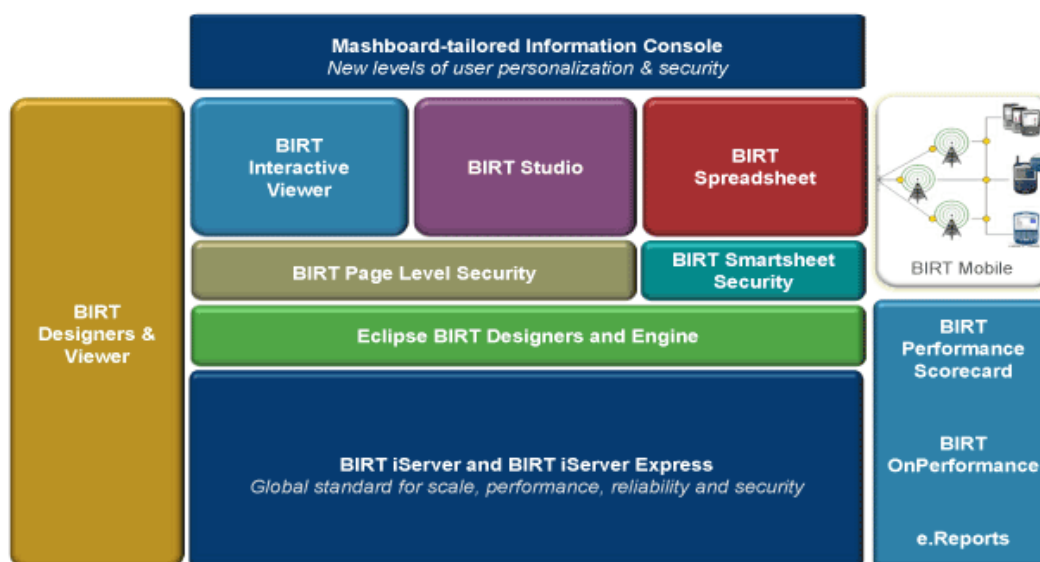
Para além da limitação de não ter dashboards nem análises interativas, a restante integração dos componentes disponíveis na edição Comunidade está menos conseguida e apresenta menor nível de funcionalidade do que a evidenciada pela solução da Pentaho.

▪ Eclipse BIRT

A plataforma enterprise do BIRT, comercializada pela Actuate, é comumente vista como uma das mais completas soluções de BI open source (lidera tanto na avaliação Magic Quadrant da Gartner 2013, como na grelha da Forrester Wave 2010). Todavia, e ainda de forma mais notória de que é o caso com a suite da Jaspersoft, a solução disponibilizada pela comunidade open source da Eclipse Foundation revela uma grande falta de componentes essenciais na sua versão de custo zero. Trata-se mesmo de apenas disponibilizar pouco mais de que componentes de reporting: essencialmente, permitindo desenhar relatórios e montar um servidor aplicacional que

os permite distribuir. Todas as restantes funcionalidades de BI não fazem parte desta versão, e teriam que ser obtidas, configuradas e mantidas de forma independente.

Figura 6 - Arquitetura funcional da plataforma comercial Actuate BIRT (fonte: InfoDecisionnel)



Assim sendo, torna-se claro que esta plataforma não permitiria à SPMS assegurar as suas necessidades de Business Intelligence para os seus projetos futuros.

▪ SpagoBI

Esta plataforma apresenta-se com uma filosofia bastante interessante, dado que apenas disponibiliza a versão Comunidade - ou seja, uma versão sem qualquer custo de aquisição ou licenciamento. Deste modo, o valor acrescentado que o sponsor desta plataforma (a empresa italiana Engineering Group) expõe ao mercado e de onde derivam as suas receitas resulta exclusivamente de serviços de treino/formação, consultoria e integração de sistemas.

Figura 7 - Arquitetura funcional da plataforma SpagoBI (fonte: SpagoBI)



A plataforma de BI em si ainda segue, em larga medida, uma abordagem de integração de diversos componentes open source fornecidos por terceiros, com a

solução SpagoBI a facultar mecanismos de integração. Recentemente, esta plataforma tem começado a desenvolver alguns componentes in-house, como, por exemplo, um motor para KPI, motor de queries guiadas (query-by-example), entre outras.

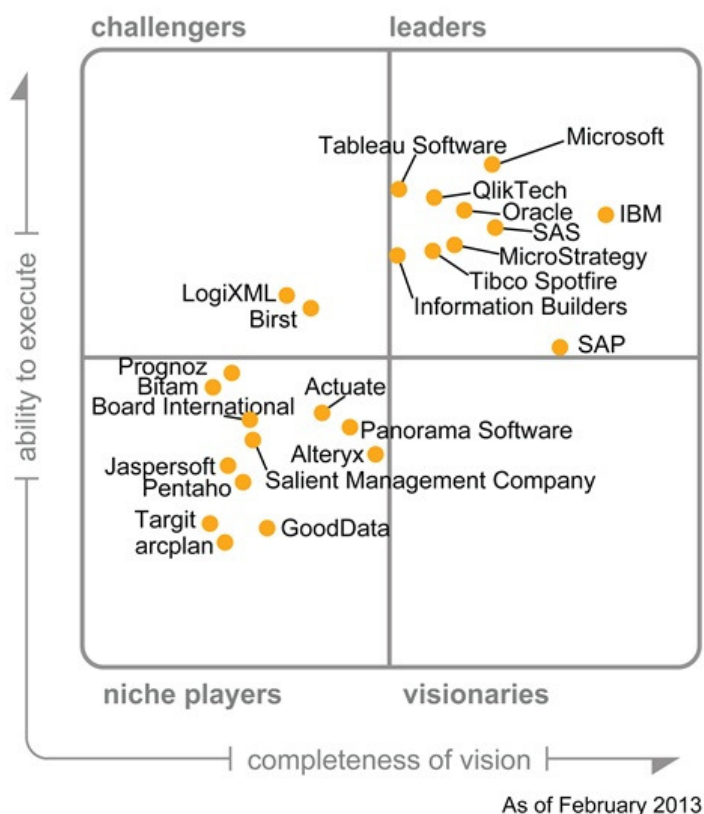
Infelizmente, esta plataforma ainda não tem um nível de disseminação muito grande, pelo que se torna difícil encontrar opiniões fundamentadas sobre as suas reais capacidades operacionais. A Gartner, por exemplo, apenas mencionou esta plataforma pela primeira vez no seu relatório de 2012 - e apenas na seção final "outros fornecedores a considerar", não fazendo parte do Quadrante Mágico. Assim, fica apenas como um candidato a seguir com atenção, dado o seu potencial.

2.2.6. Avaliação Global por Entidades Independentes

▪ Gartner Magic Quadrant Q1 2013

Em termos de posicionamento relativo na avaliação da Gartner, para o ano de 2013, a plataforma SpagoBI é a única dos 4 candidatos não representada no Quadrante Mágico. Todavia, é importante referir que as versões das 3 plataformas incluídas - Jaspersoft, Pentaho e BIRT (Actuate) - representam as versões comerciais destas plataformas de BI (e não as versões Comunidade, avaliadas neste documento):

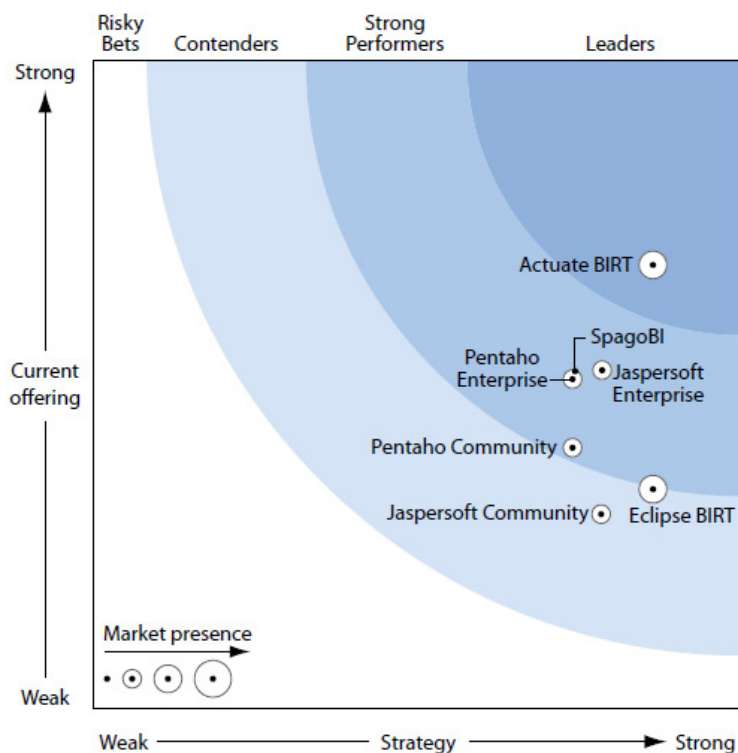
Figura 8 - "Quadrante Mágico" da Gartner para plataformas de BI - fev. 2013 (fonte: Gartner)



▪ The Forrester Wave: Open Source Business Intelligence, Q3 2010

Desta forma, talvez seja mais pertinente analisar exclusivamente as soluções open source e, adicionalmente, comparar entre versões comerciais e versões comunidade. Felizmente, a Forrester (empresa consultora de Research) publica essa mesma análise, conforme se apresenta na seguinte figura:

Figura 9 - Análise comparativa de plataformas open source de BI (fonte: Forrester Wave 3º trim. 2010)



2.2.7. Comparação Final de Plataformas de BI Open Source

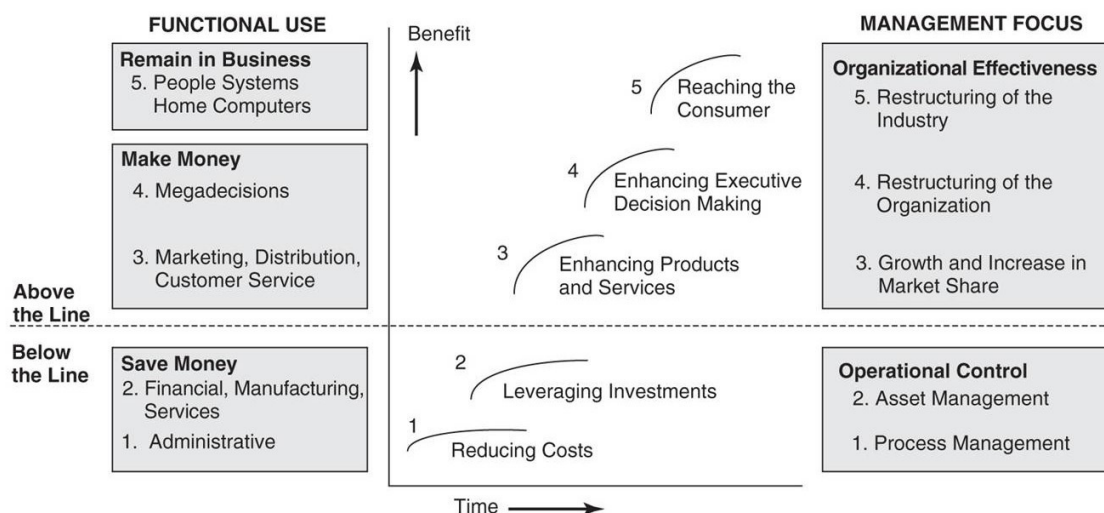
Dada a análise aqui apresentada, e tendo em conta que das quatro plataformas comparadas, uma é ainda demasiado desconhecida para se assumir como escolha segura (SpagoBI), duas revelam uma deficiência significativa de funcionalidades core nas suas versões Comunidade (Jaspersoft e Eclipse BIRT), e a restante está relativamente bem cotada e apresenta níveis interessantes de integração e funcionalidades básicas (Pentaho CE), adotaremos a plataforma open source de Business Intelligence Pentaho BI Suite na versão Community Edition (CE) para o projeto a desenvolver para a SPMS, e que será aqui documentado.

2.3. METODOLOGIAS ÁGEIS NO DESENVOLVIMENTO DE SISTEMAS DE INFORMAÇÃO

2.3.1. A Importância dos Sistemas de Informação

Os sistemas de informação têm ganho cada vez maior importância enquanto elemento integrante da estratégia das organizações - não só do ponto de vista operativo, enquanto facilitador de ganhos de eficiência e na capacidade de gerar maior volume de produção, mas também na medida em que permitem diferenciar a oferta competitiva de cada organização. Esta ideia, de que os sistemas de informação ganham cada vez maior importância, consoante vão demonstrando novas capacidades de assumir funções estratégicas nas organizações, é demonstrado pelas 5 “ondas de inovação” de Primozić (McNurlin, Sprague, & Bui, 2009):

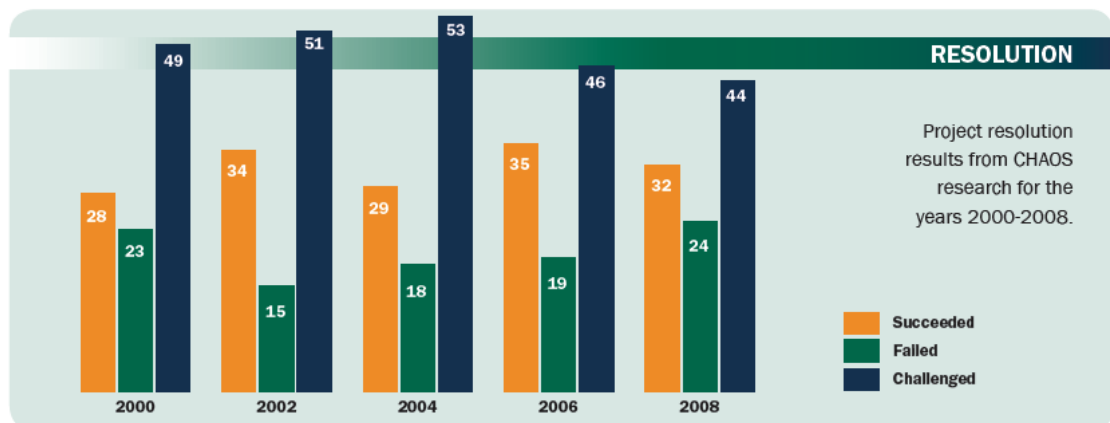
Figura 10 - As 5 Ondas de Inovação de Primozić (fonte: McNurlin, Sprague & Bui, 2009)



Source: Kenneth Primozić, Edward Primozić, and Joe Leben, *Strategic Choices: Supremacy, Survival, or Sayonara* (New York: McGraw-Hill, 1991).

Todavia, e apesar da crescente criticidade dos sistemas de informação para as organizações, continuam-se a assistir taxas de insucesso na execução de projetos SI muito marcantes. Silva & Videira salientam que as três evidências primárias do insucesso destes projetos - falta de qualidade, atraso temporal e desvio nos custos - demonstram que “estes problemas têm existido de forma contínua, e independentemente das iniciativas que têm surgido, estas não eliminaram de forma alguma o problema” (Rodrigues da Silva & Videira, 2005, p. 14). O relatório periódico produzido pelo Standish Group, intitulado “Relatório do Caos” (Chaos Report), evidencia de forma clara a elevada percentagem de projetos que não atingem um sucesso pleno:

Figura 11 - Sumário de resultados de projetos SI (fonte: Standish Group CHAOS Report 2010)



É assim evidente que taxas de sucesso absoluto em redor de apenas 30%, na última década, ressaltam a importância de rever quais as metodologias e processos que estiverem na base destes resultados, e se foram as mais adequadas.

2.3.2. Metodologias e Processos de Desenvolvimento de SI

Iremos, nesta seção, adotar o entendimento expresso por Silva & Videira dos conceitos de processo e metodologia (Rodrigues da Silva & Videira, 2005):

- **Processo:** designa uma sequência de atividades, normalmente agrupadas em fases e tarefas, que são executadas de forma sistemática, realizadas por intervenientes com responsabilidades bem definidas (nesta definição, o processo é análogo ao conceito de ciclo de vida);
- **Metodologia:** para além de pressupor a existência de um processo que lhe esteja associado, uma metodologia inclui a utilização de um conjunto de ferramentas, técnicas, notações e referências a diversos princípios e regras, com o objetivo de concretizar na prática as orientações mais teóricas do conceito de processo;

Deste modo, é possível verificar que ao longo da história do desenvolvimento de sistemas informáticos (com especial ênfase desde os anos 60), terão sido vários os processos e metodologias que foram sendo experimentadas, testadas, rejeitadas e reaproveitadas, conforme veremos na seção seguinte.

2.3.3. Evolução na utilização de metodologias e processos nos SI

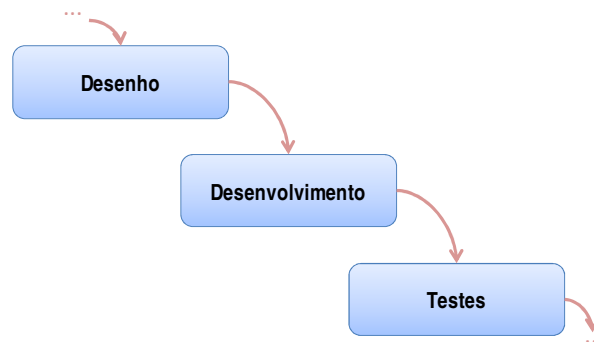
De forma sucinta, podemos resumir tanto a utilização de processos como de metodologias, em termos de grandes vagas ou fases de maior popularidade (Rodrigues da Silva & Videira, 2005):

- **A evolução nos processos de desenvolvimento de SI:**

- **Processos em Cascata (waterfall)**

Estes são considerados os modelos tradicionais de desenvolvimento de SI, em que as atividades ou fases estão sequencialmente bem delineadas, e só se avança para a etapa seguinte quando a anterior está completamente finalizada e dada como “fechada” (aceite pelo sponsor ou cliente):

Figura 12 - Exemplo de processo em cascata (fonte: própria)



Uma das principais vantagens que estes modelos preconizam é garantir às etapas seguintes um elevado nível de certeza quanto às dependências que existem entre etapas; por exemplo, garantindo-se que o desenho do sistema está terminado, a fase de desenvolvimento fica automaticamente protegida contra novos requisitos ou redesenho de funcionalidades.

Estes modelos revelam diversas fragilidades, essencialmente em redor da falta de flexibilidade que demonstram em contextos onde a mudança é prevalente (Rodrigues da Silva & Videira, 2005). Nesse sentido, foram sendo introduzidas algumas variantes, que permitem maior ou menor retroalimentação (feedback) entre etapas, como, por exemplo, o modelo em cascata revisto.

- **Processo iterativos e Incrementais**

Em resultado das dificuldades inerentes aos modelos em cascata, surgiu a ideia de realizar aperfeiçoamentos graduais ao sistema de informação, no que se convencionou chamar de modelo iterativo. Neste modelo, não se tem como objetivo atingir o estado final logo na primeira iteração de cada etapa, mas sim ir realizando várias iterações, que percorrem todo o âmbito do projeto, mas sempre com o fim de melhorar gradualmente o desenvolvimento do sistema.

No caso do modelo incremental, é o âmbito do projeto que vai sendo gradualmente revisto e alargado, até que possa atingir o objetivo global do sistema sob desenvolvimento. Nestes casos, é habitual começar com um âmbito bastante reduzido, não sendo mais de uma prova de conceito para demonstrar a exequibilidade do projeto, em termos de ferramentas, funcionalidades, contexto, equipa e outros fatores.

Todavia, o mais natural é encontrar uma conjugação destes dois modelos, no chamado processo iterativo e incremental, em que não só o sistema sob desenvolvimento é alvo de construção gradual, como o próprio âmbito do sistema (ou seja, as suas funcionalidades) são também incrementadas progressivamente.

- **A evolução nas metodologias de desenvolvimento de SI:**

- **Desenvolvimento Ad Hoc**

Inicialmente, os projetos de desenvolvimento de sistemas de informação não dispunham de qualquer metodologia formal, pelo que o sucesso dos mesmos apenas dependia da competência e conhecimento técnico dos seus intervenientes. Numa fase em que as limitações técnicas eram determinantes, a atenção estava centrada na programação e em ultrapassar limitações tecnológicas. Com o evoluir da infraestrutura de hardware, e o aumento da complexidade dos sistemas sob desenvolvimento, as desvantagens de não haver uma abordagem metodológica para estes projetos tornou-se muito evidente. Como refere Fitzgerald, esta era ficaria conhecida como a “crise de software”, com o tempo de desenvolvimento de novos sistemas a ser demasiado longo, com custos acima do expectável e que não funcionavam adequadamente (Fitzgerald, 1996).

- **Metodologias Estruturadas**

As metodologias estruturadas resultariam da “crise de software” da década de 1960, substituindo o desenvolvimento ad hoc por métodos formais, que delineavam como deviam ser realizadas as atividades subjacentes ao processo de construção dos sistemas de informação.

É possível discriminar duas grandes linhas de metodologias estruturadas, sendo que uma está mais orientada para a decomposição funcional do sistema, a que costuma designar de *análise funcional*, enquanto a outra

assenta mais na identificação dos conceitos e dados que fluem pelo sistema, a que se costuma chamar de *análise orgânica*.

São utilizadas correntemente uma grande variedade de metodologias estruturadas, até porque é habitual que projetos sob o sponsorship de grandes organizações (especialmente no caso do setor público) sejam obrigados à utilização de uma destas metodologias. Algumas das mais conhecidas serão o SSDAM (Structured System Analysis and Design Methodology), o SADT (Structured Analysis and Design Technique), o método Yourdon (Yourdon Systems Methods), e o Information Engineering, entre outros (Wieringa, 1998).

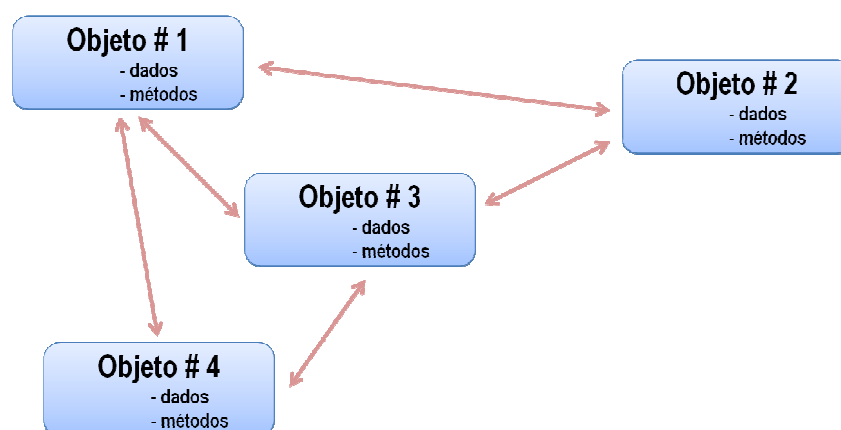
○ **Metodologias Orientadas por Objetos**

Sendo claramente um avanço sobre a fase anterior, ainda assim as metodologias estruturadas continuariam a apresentar problemas, entre os quais os seguintes (Rodrigues da Silva & Videira, 2005):

- i. Não lidam bem com sistemas de elevada complexidade;
- ii. Não estão orientadas para a atividade de manutenção dos SI;
- iii. Mantêm a dificuldade de lidar com alterações de requisitos;
- iv. A integração e reutilização de componentes e/ou código é fraca;

Para tentar resolver estas dificuldades, surgiriam as metodologias organizadas em redor do conceito de orientação por objetos:

Figura 13 - Exemplificação do conceito de "Orientação Por Objetos" (fonte: própria)



Algumas das metodologias orientadas por objetos mais conhecidas serão o Método de Booch, OMT (Object Modelling Technique), Coad-Yourdon, SOMA (Semantic Object Modelling Approach), UML (Unified Modelling Language), entre outras (Wieringa, 1998).

2.3.4. As Metodologias Ágeis

A evolução de metodologias estruturadas e orientadas por objetos trouxeram melhorias ao desenvolvimento de sistemas de informação. Ainda assim, continuaram-se a verificar casos onde projetos falharam, mesmo após adotarem essas metodologias, enquanto noutros casos a documentação formal acabaria por ocupar tanto tempo e esforço que os projetos eram terminados sem nenhum código implementado; noutros casos, projetos ficavam com pouco tempo para a fase de implementação e testes, acabando por implementar sistemas que não estavam de acordo com a documentação formal produzida no âmbito da respetiva metodologia (Lindstrom, 2004).

Em resposta a estas preocupações, foram-se tornando conhecidos casos de desenvolvimento de sistemas que se socorriam de metodologias bastante mais informais, mais ligeiras em termos de documentação e produção de artefactos, e que ainda assim atingiam graus de sucesso muito relevantes. Estas novas metodologias defendiam que “a principal documentação de um sistema é, ou devia ser, o código fonte da aplicação desenvolvida” (Rodrigues da Silva & Videira, 2005, p. 83). Alguns dos seus praticantes começaram a publicar artigos sobre as mesmas, passando algumas a ser conhecidas pela comunidade em geral: eXtreme Programming (XP), Scrum, Crystal, entre outras (Lindstrom, 2004, p. 43). Toda esta vaga de fundo sobre o agilizar das metodologias tradicionais culminaria num retiro, realizado em 2001 e onde participariam os principais inovadores e autores destas novas abordagens, de onde saiu o Manifesto Agile para o Desenvolvimento de Software:

Tabela 3 - O Manifesto para Desenvolvimento Ágil de Software (fonte: agilemanifesto.org)

Ao desenvolver e ao ajudar outros a desenvolver software, temos vindo a descobrir melhores formas de o fazer.		
Através deste processo começámos a valorizar:		
Indivíduos e interações mais do que processos e ferramentas		
Software funcional mais do que documentação abrangente		
Colaboração com o cliente mais do que negociação contratual		
Responder à mudança mais do que seguir um plano		
<i>Ou seja, apesar de reconhecermos valor nos itens à direita, valorizamos mais os itens à esquerda.</i>		
Kent Beck	James Grenning	Robert C. Martin
Mike Beedle	Jim Highsmith	Steve Mellor
Arie van Bennekum	Andrew Hunt	Ken Schwaber
Alistair Cockburn	Ron Jeffries	Jeff Sutherland
Ward Cunningham	Jon Kern	Dave Thomas
Martin Fowler	Brian Marick	

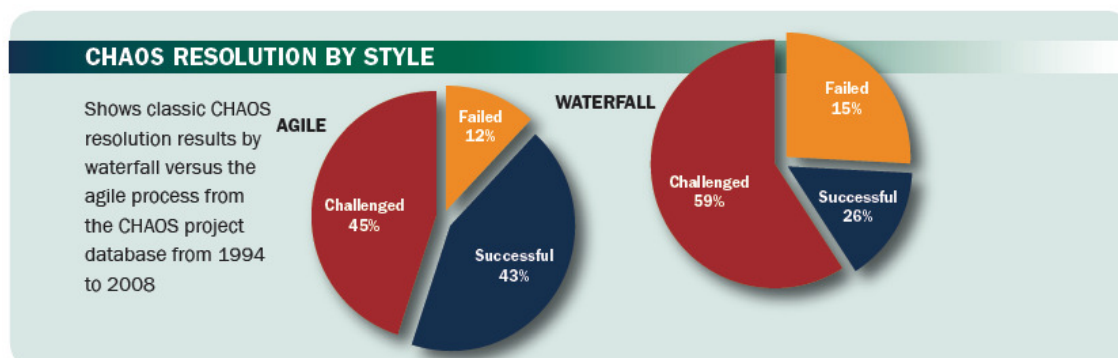
O Manifesto tem, como parte integrante, 12 Princípios refletindo o que deve ser praticado nas metodologias Ágeis de desenvolvimento de SI:

Tabela 4 - Os 12 Princípios do Manifesto Ágil (fonte: agilemanifesto.org)

1. A nossa maior prioridade é, desde as primeiras etapas do projeto, satisfazer o cliente através da entrega rápida e contínua de software com valor.
2. Aceitar alterações de requisitos, mesmo numa fase tardia do ciclo de desenvolvimento. Os processos ágeis potenciam a mudança em benefício da vantagem competitiva do cliente.
3. Fornecer frequentemente software funcional. Os períodos de entrega devem ser de poucas semanas a poucos meses, dando preferência a períodos mais curtos.
4. O cliente e a equipa de desenvolvimento devem trabalhar juntos, diariamente, durante o decorrer do projeto.
5. Desenvolver projetos com base em indivíduos motivados, dando-lhes o ambiente e o apoio de que necessitam, confiando que irão cumprir os objetivos.
6. O método mais eficiente e eficaz de passar informação para e dentro de uma equipa de desenvolvimento é através de conversa pessoal e direta.
7. A principal medida de progresso é a entrega de software funcional.
8. Os processos ágeis promovem o desenvolvimento sustentável. Os promotores, a equipa e os utilizadores deverão ser capazes de manter, indefinidamente, um ritmo constante.
9. A atenção permanente à excelência técnica e um bom desenho da solução aumentam a agilidade.
10. Simplicidade – a arte de maximizar a quantidade de trabalho que não é feito – é essencial.
11. As melhores arquiteturas, requisitos e desenhos surgem de equipas auto-organizadas.
12. A equipa reflete regularmente sobre o modo de se tornar mais eficaz, fazendo os ajustes e adaptações necessárias.

De facto, há alguma razão para acreditar que metodologias Ágeis podem, em muitos casos, revelar melhores resultados do que as metodologias tradicionais, especialmente em contexto de mudanças imprevisíveis (Nerur, Mahapatra, & Mangalaraj, 2005):

Figura 14 – Comparação entre metodologias tradicionais e Ágeis (fonte: Standish Group CHAOS Report 2010)



3. METODOLOGIA

Nesta secção do relatório de projeto, iremos detalhar os diversos aspetos metodológicos que estiveram subjacentes ao desenvolvimento e implementação do projeto de business intelligence aqui documentado. Deste modo, serão referidos não só aspetos contextuais (organizacionais, funcionais e técnicos) mas também algumas das opções tomadas quanto à execução dos processos de desenvolvimento do projeto.

3.1. ENQUADRAMENTO ORGANIZACIONAL: A SPMS, EPE

O projeto aqui apresentado é realizado sob a égide da SPMS – Serviços Partilhados do Ministério da Saúde, EPE, pelo que importa enquadrar o papel desta organização face ao tema deste trabalho.

A SPMS é uma Entidade Pública Empresarial (EPE) com a responsabilidade de assegurar a prestação de *serviços partilhados* ao nível de compras e logística, gestão financeira, recursos humanos especializados e sistemas de informação para as entidades que integram o Serviço Nacional de Saúde (SNS).

A criação da SPMS foi aprovada em reunião de Conselho Ministros de 17 de dezembro de 2009 e a sua formalização foi consagrada no Decreto-Lei nº 19/2010, publicado a 22 de Março de 2010. A SPMS é uma pessoa coletiva de direito público empresarial, gozando de autonomia financeira, administrativa e empresarial, com supervisão e controlo dos ministérios da Saúde e Finanças.

3.1.1. Missão

A Missão da SPMS é a seguinte (Decreto-Lei n.º 19/2010, de 22 de março):

“Prestação de serviços partilhados específicos da área da saúde em matéria de compras e logística, gestão financeira, recursos humanos e sistemas de informação e comunicação e demais atividades complementares e subsidiárias às entidades que integram o Serviço Nacional de Saúde (SNS), independentemente da sua natureza jurídica, bem como aos órgãos e serviços do Ministério da Saúde e a quaisquer outras entidades, quando executem atividades específicas da área da saúde.”

Enquadram-se na missão da SPMS as seguintes atribuições principais (Decreto-Lei n.º 19/2010, de 22 de março):

“a) Cooperação com os serviços de saúde na inovação e gestão da mudança;

b) Promover partilha de conhecimento para fortalecimento da reforma do sector da saúde;

c) *Prestação de serviços partilhados em diversas áreas como a gestão financeira, contabilidade, recursos humanos de elevada eficiência, serviços de compras e logística e sistemas e tecnologias de informação e comunicação, atentas as necessidades e responsabilidades dos serviços da rede SNS;*

d) *Funcionamento como central de compras na aquisição de bens e serviços para o sector saúde, obtendo-se ganhos económicos pelo efeito de escala e negociação em diálogo concorrencial com os fornecedores;*

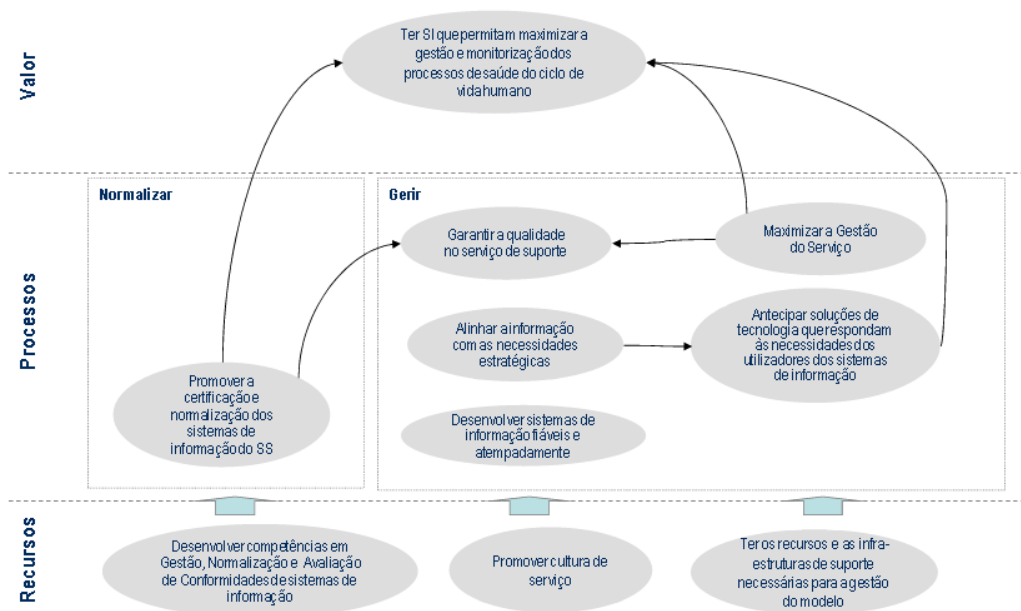
e) *Contribuir como Unidade Operativa do Ministério da Saúde para a promoção da modernização e inovação nos procedimentos operacionais dos Hospitais e Agrupamentos de Centros de Saúde, pela adequada integração de tecnologias de informação e comunicação, em perfeita articulação com as políticas nacionais e regionais.”*

3.1.2. Responsabilidade sobre os sistemas TIC do MS/SNS

Em Novembro de 2011, transitaram para a SPMS as atribuições quanto à prestação de serviços partilhados em termos de sistemas e tecnologias de informação e comunicação das entidades do SNS e dos órgãos e serviços do Ministério da Saúde, através do Decreto-Lei nº 108/2011, de 17 de novembro.

Estas atribuições eram anteriormente da responsabilidade da Administração Central do Sistema de Saúde, IP (ACSS), que passa agora a contar com a SPMS para “prover o Serviço Nacional de Saúde com os adequados sistemas de informação e comunicação e mecanismos de racionalização de compras” (Decreto-Lei n.º 108/2011, de 17 de novembro). Deste modo, passa a SPMS a assumir a responsabilidade de concretizar a estratégia delineada pela ACSS para esta área de intervenção (ACSS, 2010):

Figura 15 - Mapa de Intervenção Estratégica para a Área de SI (fonte: ACSS 2010)



Deste modo, e no que concerne especificamente sistemas de informação, a SPMS tem como responsabilidade (Decreto-Lei n.º 108/2011, de 17 de novembro):

“... a cooperação, a partilha de conhecimentos e informação e o desenvolvimento de atividades de prestação de serviços nas áreas dos sistemas e tecnologias de informação e de comunicação, garantindo a operacionalidade e segurança das infraestruturas tecnológicas e dos sistemas de informação do Ministério da Saúde e promovendo a definição e utilização de normas, metodologias e requisitos que garantam a interoperabilidade e interconexão dos sistemas de informação da saúde, entre si e com os sistemas de informação transversais à Administração Pública.”

3.1.3. A Lógica de Funcionamento da SPMS

Conforme refere o Decreto-Lei nº 19/2010, de 22 de março:

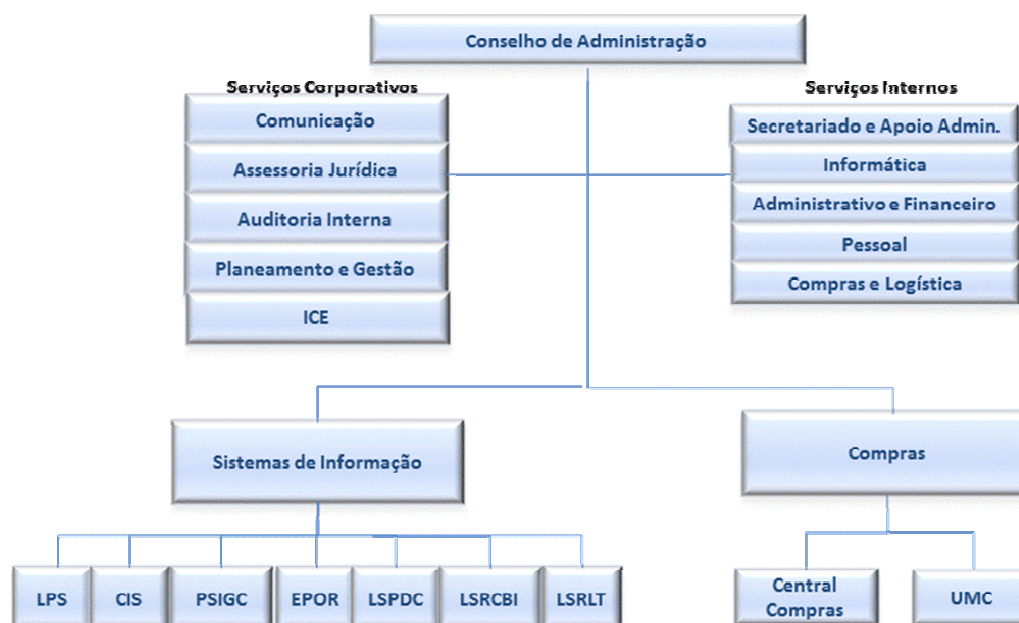
“Deixa de pender sobre o Estado e as instituições e serviços do SNS a responsabilidade pelo fornecimento de serviços que, pela sua natureza e relevância estratégica, podem e devem ser garantidos por um único fornecedor a todas as entidades do sistema de saúde, o que permite a sua libertação para se concentrarem na prossecução das suas atividades nucleares: a prestação de cuidados de saúde aos cidadãos.”

A SPMS, EPE pode assim estabelecer acordos-quadro, bem como realizar Concursos Públicos de Aprovisionamento (CAP) com fornecedores de produtos e serviços, que poderão ser posteriormente adquiridos e consumidos pelas entidades de saúde do SNS que escolham aderir aos mesmos. O facto de a negociação ser central e com quantidades superiores, permite obter reduções de preço e condições contratuais privilegiadas, potenciando os ganhos resultantes das economias de escala.

3.1.4. Organização da SPMS

A SPMS está dividida em duas áreas principais de prestação de serviços: os Sistemas de Informação e a área de Compras; conta ainda com duas áreas de serviços de natureza interna: os Serviços Corporativos e os Serviços Internos, conforme evidenciado no respetivo organograma (SPMS, 2013):

Figura 16 - Organograma da SPMS (fonte: SPMS)



3.1.5. Linha de Serviços de Registos Centrais e Business Intelligence

A área de prestação de serviços de Sistemas de Informação está, por sua vez, subdividida em 7 áreas de coordenação, sendo que uma delas respeita especificamente aos sistemas cuja natureza é de registo central (entenda-se nacional ou transversal ao SNS) e de business intelligence. Em termos de relevância para o projeto que aqui se apresenta, importa referir que uma das competências desta Linha de Serviço é “definir uma política de Business Intelligence para o Ministério da Saúde e garantir a implementação efetiva de soluções no curto e médio prazo” (SPMS, 2013). É na procura de respostas a esta competência que se insere o trabalho aqui apresentado, no sentido de explorar metodologias, ferramentas e práticas que possibilitem maior eficácia e eficiência na implementação de soluções de business intelligence.

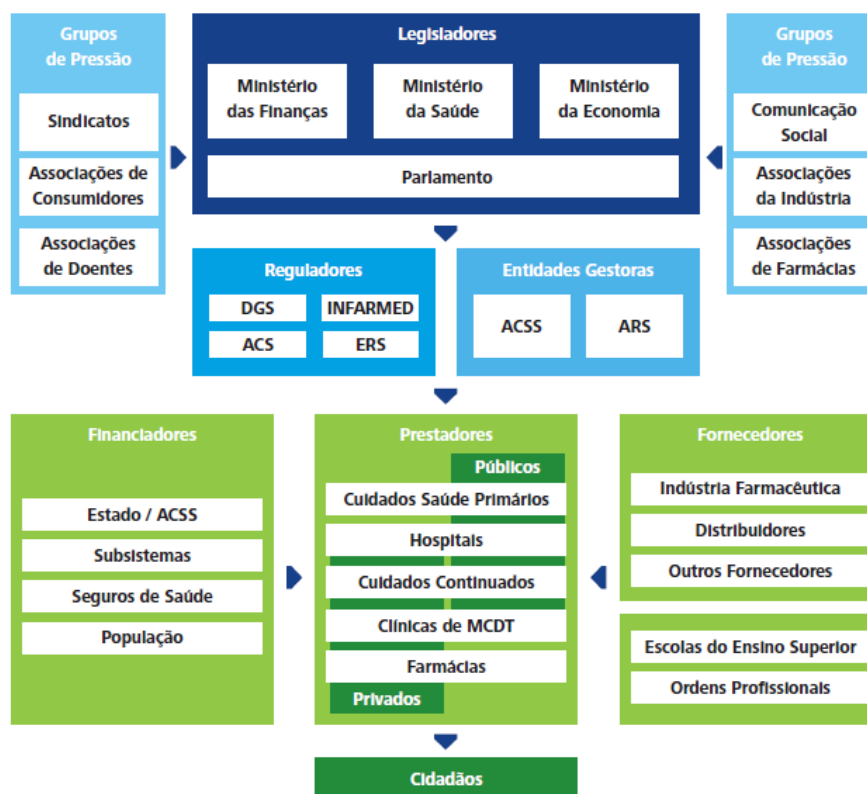
3.2. ENQUADRAMENTO FUNCIONAL: A GESTÃO DE PROJETOS

3.2.1. Enquadramento Setorial da Gestão de SI pela SPMS

A SPMS é contratada diretamente pela ACSS, através da figura de um Contrato-Programa, para prover o SNS com os adequados sistema e tecnologias de informação e comunicação. Para o ano de 2012, por exemplo, o valor estabelecido como contrapartida financeira pela prestação deste serviço foi de 42,3 milhões de euros (SPMS, 2012). Cabe, assim, à SPMS organizar e gerir uma resposta transversal em

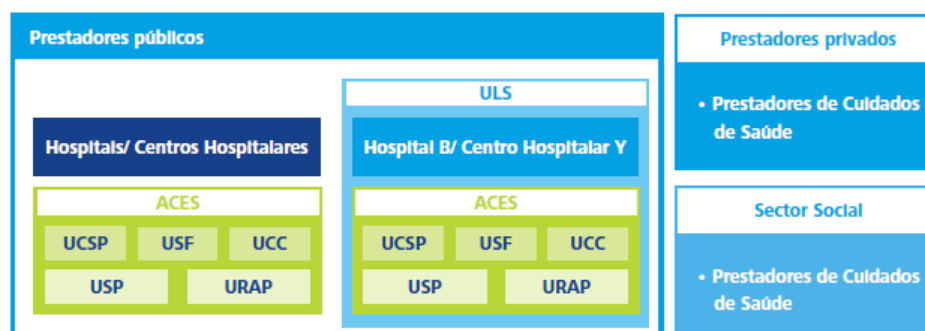
termos de SI para as entidades do SNS, mas tendo obrigatoriamente em conta a multiplicidade de stakeholders adstritos ao setor da saúde português, conforme se vê na Figura 17:

Figura 17 - Principais stakeholders no sector da saúde em Portugal (fonte: Deloitte 2011)



A área de atuação da SPMS, no que concerne a disponibilização de sistemas, infraestrutura e soluções, tem uma maior preponderância sobre as entidades prestadoras de cuidados de saúde, tanto em termos da rede nacional hospitalar como de cuidados de saúde primários:

Figura 18 - Prestadores de cuidados de saúde em Portugal (fonte: Deloitte 2011)



No entanto, os projetos da SPMS, na área de SI, não se esgotam nos prestadores de cuidados de saúde. A SPMS atua transversalmente por todo o setor da saúde, em termos de portais, sistemas e plataformas centralizadas de dados e informação.

3.2.2. Os Projetos SI na SPMS

Atualmente, a SPMS encontra-se envolvida no desenvolvimento de quase meia centena de projetos nas áreas de Cuidados Primários e Cuidados Hospitalares, destacando-se na Figura 19 os que foram publicamente referidos na reunião de Trabalho sobre Tecnologias de Informação e Comunicação (TIC), promovida pela SPMS, EPE no dia 11 de julho de 2013 (SPMS, 2013):

Figura 19 - Amostra de projetos de SI da SPMS (fonte: SPMS 2013)

PARA OS HOSPITAIS

PRESCRIÇÃO

- . A PEM substituirá a prescrição no SAM até final de agosto de 2013
- . A PEM será implementada nos hospitais sem SAM nos meses de setembro e outubro de 2013

SICO

- . O Rollout do SICO terminará em novembro de 2013

WebRNU

- . O Rollout do WebRNU nos hospitais estará concluído no final de agosto de 2013
- . Nos Hospitais com maternidade estará também disponível o NascerUte

SONHO

- . O SONHO será uma plataforma administrativa comum aos Hospitais e Cuidados de Saúde Primários

SIDC

- . Será o software agregador da contabilidade

RHV

- . Centralização concluída em todos os Hospitais até final de Novembro de 2013

SIGLIC e CTH

- . Estão a ser tomadas as ações necessárias para implementar as melhorias/correções indispensáveis e para proceder à internalização destes sistemas

PDS

- . É importante a divulgação interna da PDS de modo a incentivar a sua utilização
- . Está em curso a sua implementação nas regiões autónomas

SEGURANÇA

- . Foram apresentados os resultados preliminares do survey nacional
- . Está em curso a criação de uma nova política de segurança

INFRAESTRUTURAS

- . Está em curso o assessment para a implementação do SONHO
- . Está a decorrer o diálogo concorrencial para a Rede Informática da Saúde

CENTRO DE SUPORTE

- . PEM e SICO terão suporte 24h

SISTEMA DE GESTÃO DE TRANSPORTE DE DOENTES NÃO URGENTES

- . Processo central de aquisição para os hospitais da Aplicação de Gestão Integrada de Transportes (AGIT)

SVIDA

- . Importante garantir o preenchimento dos dados e acompanhamento da situação

HARDWARE

- . Aquisição centralizada.
- . Investimento responsável

SOBRE CUIDADOS DE SAÚDE PRIMÁRIOS

PRESCRIÇÃO

- . PEM - Atualização dos postos de trabalho durante os meses de setembro e outubro

CENTRO DE SUPORTE

- . PEM e SICO terão suporte 24h

WebRNU

- . ARS Norte - está em conclusão

SONHO

- . O SONHO será uma plataforma administrativa comum aos Hospitais e Cuidados de Saúde Primários

SIDC

- . ARS Norte será a primeira região onde será feito o piloto do upgrade - SIDCv2

PDS

- . É importante incentivar a utilização do Portal do Utente

SEGURANÇA

- . Foram apresentados os resultados preliminares do survey nacional
- . Está em curso a criação de uma nova política de segurança

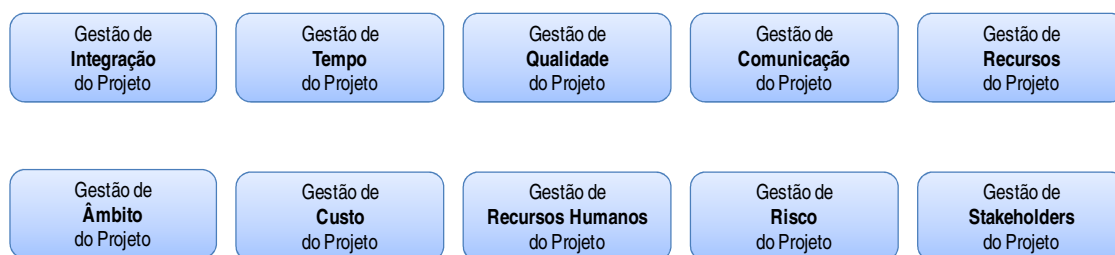
3.2.3. Gestão de Projetos SI na SPMS: Fundamento para o BI-GEST

A SPMS gere os projetos de SI com base nos fundamentos do guia Project Management Body of Knowledge (PMBOK). Este guia é produzido pelo Project Management Institute (PMI), e preconiza uma gestão de projetos baseada em processos. O Guia PMBOK baseia-se em 5 processos transversais a todo o tipo de projetos:

- i. Iniciação (Initiating)
- ii. Planeamento (Planning)
- iii. Execução (Executing)
- iv. Monitorização e Controlo (Monitoring and Controlling)
- v. Encerramento (Closing)

e em 10 áreas de conhecimento, que se cruzam com estes processos:

Figura 20 - As 10 Áreas de Conhecimento do Guia PMBOK (fonte: PMBOK)



A SPMS detém vários sistemas transacionais que apoiam algumas destas áreas de conhecimento, como são o caso, por exemplo, do sistema de Time Report (tempo), do sistema financeiro (custo), do sistema Redmine (qualidade|risco|âmbito), entre outros. Todavia, nenhum destes sistemas operacionais está interconectado, nem está disponível qualquer mecanismo de exploração de informação consolidada e centrada no projeto.

Por outras palavras, a SPMS não beneficia de nenhum sistema de business intelligence ou de analytics para a gestão dos seus projetos de SI. Neste momento, o acompanhamento estratégico destes projetos é efetuado através da compilação ad hoc de informação oriunda destes diversos sistemas, para depois ser trabalhada e apresentada com recursos a folhas de cálculo ou documentos isolados.

Desta forma, parece-nos claro a necessidade de ser desenvolvido e implementado um sistema de business intelligence para a gestão de projetos SI da SPMS. A esse futuro sistema damos o nome de BI-GEST, que iremos detalhar melhor na restante parte do trabalho aqui apresentado.

3.3. ASPETOS METODOLÓGICOS DO PROJETO BI-GEST

Apresentaremos, de seguida, alguns aspetos relevantes em termos de conceitos e práticas seguidas na concretização do trabalho aqui apresentado.

3.3.1. Âmbito

O âmbito do projeto descrito neste trabalho, de implementação do sistema “BI-GEST” para a área de gestão de projetos na SPMS, apresenta-se reduzido face às necessidades reais da organização.

Como vimos na seção anterior, caso se considerasse a implementação de um data warehouse corporativo para a SPMS (note-se que não se trata de DW para o SNS, tarefa imensuravelmente mais complexa), seria necessário considerar todas as 10 áreas de conhecimento referidas pelo PMBOK e, ainda, áreas adicionais que digam respeito à atividade da Central de Compras.

Neste caso concreto, e face a limitações de várias ordens – entre as quais o prazo restrito, a indisponibilidade de diversos recursos e outras limitações organizativas – apenas iremos desenvolver uma data mart para a gestão do tempo afeto aos diversos projetos e respetivas atividades. Neste momento, o sistema transacional que se ocupa destes registos apenas disponibiliza dados de colaboradores da SPMS, pelo que ficam de fora do âmbito deste projeto os tempos associados aos recursos de empresas subcontratadas.

Não tendo ainda sido possível assegurar um acesso contínuo ao sistema financeiro, iremos incorporar o elemento de custo do projeto através de uma métrica calculada (taxa horária vezes n.º de horas do projeto) enquanto proxy do custo real dos recursos afetos a cada projeto.

3.3.2. Processo

O processo de desenvolvimento deste projeto será iterativo e incremental, conforme detalhado na seção 2.3.3 deste documento.

3.3.3. Metodologia

A metodologia que será seguida na concretização do projeto BI-GEST é de natureza Ágil (descrita na seção 2.3.4), sendo que não nos iremos orientar por nenhuma framework específica (como, por exemplo, seguindo práticas de Scrum). Todavia, as premissas e princípios do Manifesto Ágil irão, em última análise, nortear todo o desenvolvimento deste projeto.

3.3.4. Infraestrutura

Dada a parca exigência tecnológica de todos os produtos utilizados no âmbito deste projeto, o ambiente de desenvolvimento do mesmo será viabilizado através da utilização de PCs de secretária (vulgo desktop). Todavia, e mantendo a inspiração das metodologias Ágeis, iremos, desde o início do projeto, experimentar com o alojamento do mesmo em ambiente Cloud – concretamente, recorrendo à infraestrutura da Amazon Web Services (AWS), por via dos serviços de Elastic Compute Cloud (EC2) e Relational Database Service (RDS). Considerando as necessidades de segurança para alojamento de dados na cloud, a versão do BI-GEST alojada na cloud da Amazon conta com dados anonimizados (nomes de recursos humanos e nomes de projetos, sendo os valores quantitativos reais).

Quando da passagem a produção, será decidida qual a infraestrutura mais adequada, tendo em conta as disponibilidades do data center da SPMS e o nível de utilização que esteja previsto.

3.3.5. Software

Os aspetos referentes à escolha de software para este projeto poderão ser melhor compreendidos, separando as necessidades de base de dados e armazenamento (necessidades de data warehouse / data mart), das referentes à plataforma de reporting e análise (necessidades de servidor BI).

No que concerne a camada de armazenamento da dados, e a escolha por uma sistema de gestão de base de dados (SGBD), importa referir que a preocupação de que seja open source não é, neste caso, crítica. A SPMS está neste momento a estudar a hipótese de normalizar e recomendar que esta camada seja exclusivamente assegurada, de forma transversal em todos os projetos, por uma solução proprietária (concretamente, com a solução Oracle). Deste modo, podíamos ter avançado com uma solução Oracle também para este projeto. No entanto, e sabendo que isso não cria qualquer incompatibilidade ou inconveniência numa futura migração do sistema de gestão de base de dados, decidimos avançar com a utilização do sistema de gestão de base de dados open source MySQL, na sua versão 5.6. Dado que será provável a migração mais tarde para sistema proprietário Oracle, escolhemos utilizar exclusivamente o motor de storage InnoDB na base de dados MySQL utilizada neste projeto, por ser o que mais se aproxima do funcionamento da base de dados Oracle.

Em termos de camada de reporting e análise de dados, era necessário assegurar a escolha de uma solução exclusivamente open source. Dada a análise comparativa

realizada na seção 2.2.7 deste relatório, escolhemos recorrer à plataforma Pentaho Community Edition, bem como aos respectivos plug-ins open source que estão disponíveis para a mesma (nomeadamente, C-Tools e Saiku).

4. IMPLEMENTAÇÃO

Conforme referido, a implementação da solução BI-GEST foi realizada com base em metodologias Agile, pelo que decorreram diversas iterações de desenvolvimento até estabelecer a solução aqui descrita. Aliás, numa lógica puramente Agile, não é possível, nem relevante, pré-determinar quantos ciclos de desenvolvimento será necessários – por definição, o desenvolvimento nunca está “fechado”, pois assenta num paradigma de evolução contínua.

Todavia, e devido à necessidade de avaliar e relatar um determinado estágio desta solução, escolheu-se considerar o BI-GEST após as primeiras 4 iterações. A melhor forma para se compreender a diferença fundamental quando falamos de desenvolvimento Agile será através da observação concreta do que se passou em cada iteração, pelo que iremos abordar sistematicamente cada um destes ciclos.

4.1. ITERAÇÃO # 0 – PROVA DE CONCEITO

A premissa base desta primeira interação seria conseguir provar que as ferramentas delineadas para o projeto funcionam, e que demonstram capacidades suficientes para assegurar os resultados pretendidos pelo cliente. Escolhendo implementar apenas uma prova de conceito, torna-se possível reduzir o projeto à sua essência mais elementar, adiando – para ciclos de desenvolvimento posteriores – decisões sobre aspetos mais complexos deste projeto.

Após deliberações conjuntas com o cliente, estabeleceu-se como meta mínima o cumprimento do seguinte objetivo:

Após recebimento de um sub-conjunto de dados-fonte (em formato real, semelhante a “dados de produção”), conseguir extrair e carregar para um data mart (concebido como parte integrante de futuro data warehouse) esses dados, permitindo obter um relatório simples diretamente da aplicação de BI.

Sendo certo que este objetivo revela que há ainda muitas facetas típicas de processo de BI que não serão investigadas (por exemplo, dashboards e análises ad hoc), permitiu estabelecer uma base inicial para 3 aspetos fundamentais: desenho do DW, ETL sobre os dados-fonte e capacidade de reporting.

4.1.1. Desenho do Data Warehouse

Na fase de desenhar a estrutura de dados que permitiria receber os dados-fontes, foi tida em conta a abrangência da informação contida no ficheiro de folha de cálculo, reunindo os dados-fonte, conforme revela a Figura 21:

Figura 21 - Formato de dados-fonte para utilização em prova de conceito (fonte: própria)

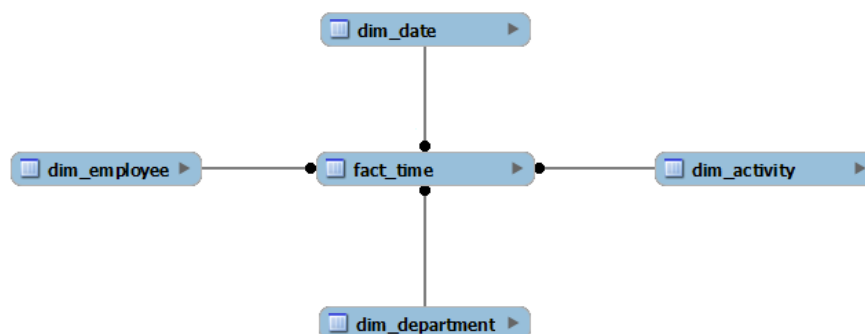
Num_Empregado	Desc_1	Desc_2	Desc_3	Desc_4	Desc_5	Total Horas	2ª feira	3ª feira	4ª feira	5ª feira	6ª feira	sábado	domingo	Periodo
234	Coord_Proj_Recursos	Desenvolv_Projetos	Famig	Gestão de Projeto		1					1			03-06-2013
244	Coordenação_Aplicações	Manutenção_Aplicações	SAM_H	Manutenção Corretiva		10	2	4		2,5	2			03-06-2013
244	Coordenação_Aplicações	Manutenção_Aplicações	SAM_H	Manutenção Corretiva		3			3					10-06-2013
244	Coordenação_Aplicações	Manutenção_Aplicações	SAM_H	Manutenção Corretiva		19	4	4	3	5	3			17-06-2013
244	Coordenação_Aplicações	Manutenção_Aplicações	SAM_H	Manutenção Corretiva		4				3	1			24-06-2013
248	Coordenação_Aplicações	Manutenção_Aplicações	SAM_CS	Manutenção Corretiva		16	3	0	6	3	4			03-06-2013
248	Coordenação_Aplicações	Manutenção_Aplicações	SAM_CS	Manutenção Corretiva		22	0	6	4	8	4			10-06-2013
248	Coordenação_Aplicações	Manutenção_Aplicações	SAM_CS	Manutenção Corretiva		25	7	6	4	6	2			17-06-2013
248	Coordenação_Aplicações	Manutenção_Aplicações	SAM_CS	Manutenção Corretiva		18	0	6	5	5	2			24-06-2013
250	Coordenação_Aplicações	Manutenção_Aplicações	SAM_H	Manutenção Corretiva		8	2	1		4	1			03-06-2013
250	Coordenação_Aplicações	Manutenção_Aplicações	SAM_H	Manutenção Corretiva		8,4		1,9	1,9	2,6	2			10-06-2013
250	Coordenação_Aplicações	Manutenção_Aplicações	SAM_H	Manutenção Corretiva		29,8	7,2	6	6,6	5	5			17-06-2013
250	Coordenação_Aplicações	Manutenção_Aplicações	SAM_H	Manutenção Corretiva		2,1		1,1	1					24-06-2013
252	Coordenação_Aplicações	Manutenção_Aplicações	RNU	Suporte 2ª Linha		14,5	3		4	3,5	4			03-06-2013
252	Coordenação_Aplicações	Manutenção_Aplicações	RNU	Suporte 2ª Linha		11		2	3,5	3	2,5			10-06-2013
252	Coordenação_Aplicações	Manutenção_Aplicações	RNU	Suporte 2ª Linha		20,5	5	3	4	3,5	5			17-06-2013
253	Coord_Proj_Recursos	Desenvolv_Projetos	SONHO2	Gestão de Projeto		37	8	4	3	10	12			03-06-2013
253	Coord_Proj_Recursos	Desenvolv_Projetos	SONHO2	Gestão de Projeto		33	3	8	2	10	10			17-06-2013

A conceção deste DW seguiu as 4 etapas recomendadas por Kimball (Kimball & Ross, 2002), nomeadamente:

1. **Escolha de Processo de Negócio** – a escolha foi natural, dado que o cliente tinha já estabelecido que, por enquanto, apenas havia acesso aos dados do Relatório de Tempos de Atividade (vulgo Time Report);
2. **Declaração da Granularidade** – tendo em conta a natureza do reporte de tempos evidenciada nos dados-fonte, estabeleceu-se a granularidade mais baixa possível: as horas realizadas em cada dia, por cada um dos colaboradores, em cada uma das atividades realizadas (considerando todas as sub-divisões de atividade);
3. **Escolha das Dimensões** – neste iteração, de prova de conceito, reduziu-se a escolha a apenas as seguintes dimensões:
 - a. Data (data diária)
 - b. Colaborador
 - c. Departamento (representa a lógica de Linha de Serviço ou Unidade Orgânica, presente na SPMS)
 - d. Atividade (nesta fase, engloba simultaneamente o conceito de Projeto e de Atividade, incluindo as sub-atividades)
4. **Identificação dos Factos** – também aqui o trabalho foi direto, dado apenas estar disponível a métrica “horas realizadas” nos dados-fonte;

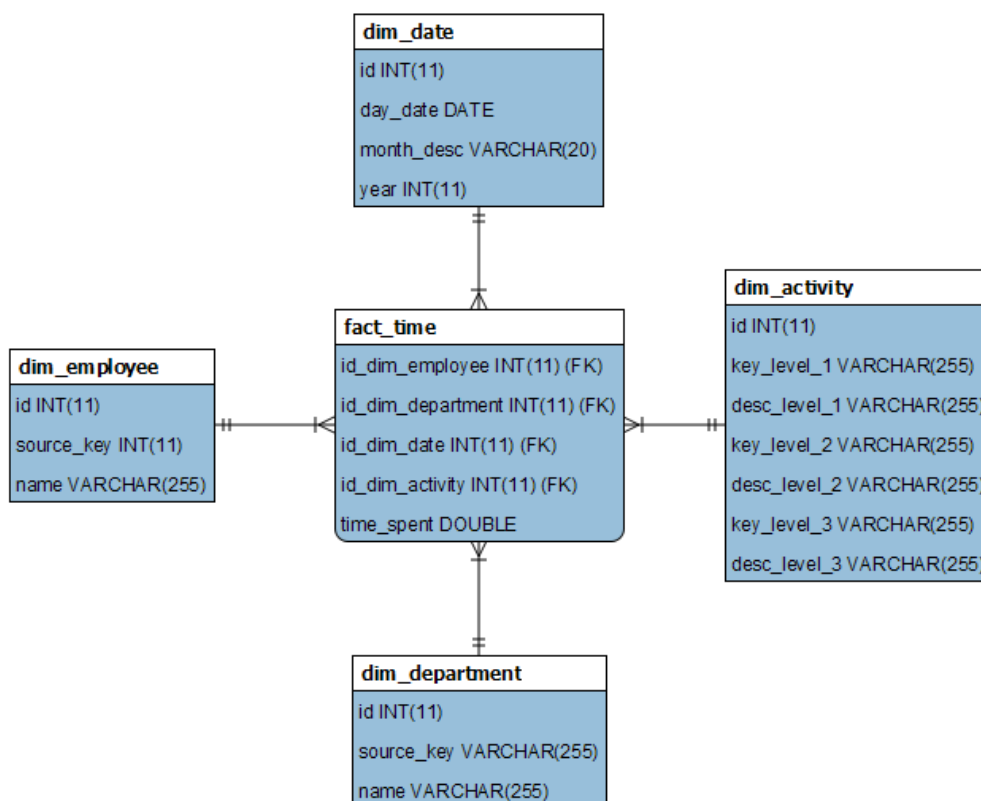
Deste modo, o modelo lógico, em forma de star schema, revela-se extremamente simples – aliás, como será recomendável numa prova de conceito – e encontra-se na seguinte figura:

Figura 22 - Modelo lógico (star schema) para a primeira iteração (fonte: própria)



A implementação deste modelo para a sua representação física, em sistema de gestão de base de dados MySQL 5.6, também não apresentou dificuldades relevantes, resultando no seguinte esquema ERD:

Figura 23 - Modelo físico do data mart na primeira iteração (fonte: própria)



Note-se que nesta iteração, foram propositadamente descuradas questões que pudessem atrasar o desenho do data mart, tais como a identificação e modelação de slowly-changing dimensions, identificação de todos os atributos das dimensões, entre

outras. Numa lógica Agile, este modelo é completo o suficiente para obter um primeiro resultado, e robusto o suficiente para permitir ser melhorado e incrementado em iterações posteriores.

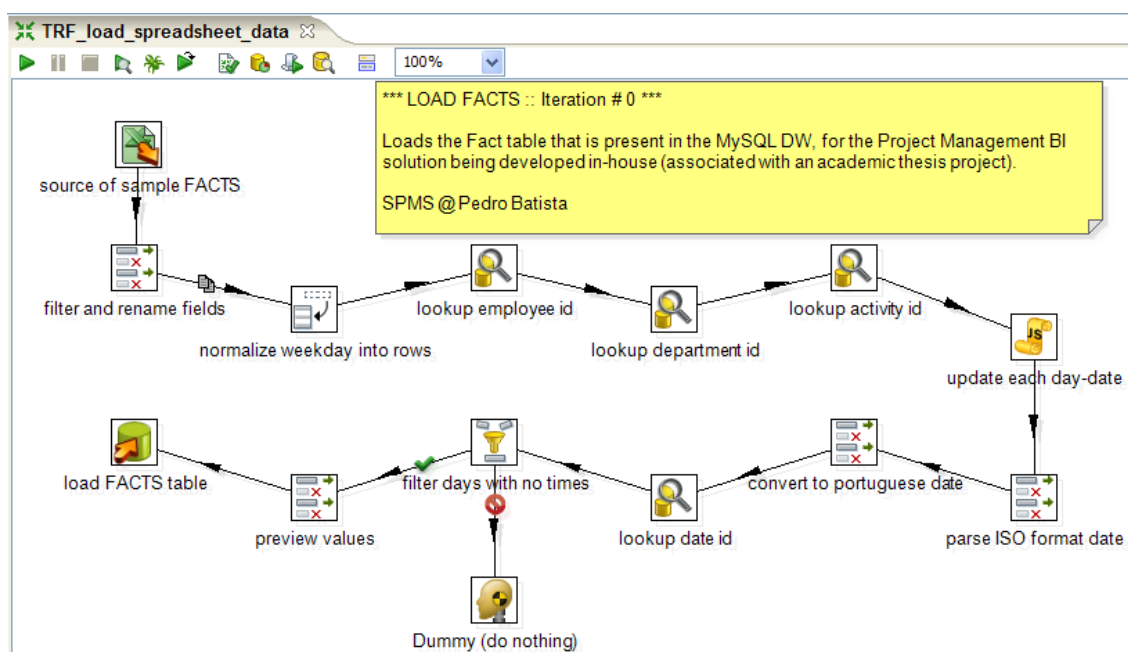
4.1.2. Processos de ETL

Durante a fase de extração, tratamento e carregamento dos dados-fonte, foram assumidas algumas concessões ao que seria um processo completo de ETL, fruto da natureza experimental e simples desta primeira iteração, nomeadamente:

- As tabelas das dimensões não foram alvo de processo ETL automatizado, dada a sua reduzida complexidade, tendo sido utilizado scripts SQL específicos para a respetiva criação e preenchimento das tabelas;
- A dimensão Data foi preenchida com dados gerados de forma autónoma, tendo esses dados sido aproveitados de outro projeto. Adicionalmente, apenas foram gerados dados suficientes para o mês em questão (datas para os 30 dias do mês de junho de 2013);
- Dada a ausência de validações e tratamento de erros sobre os dados-fonte nesta iteração, não foi utilizada uma área de staging para o ETL;

Nesta primeira iteração, o único processo de ETL que recebeu já uma implementação mais completa, com recurso à ferramenta open source de ETL Pentaho Data Integration (PDI), foi o carregamento dos factos, referenciado na Figura 24:

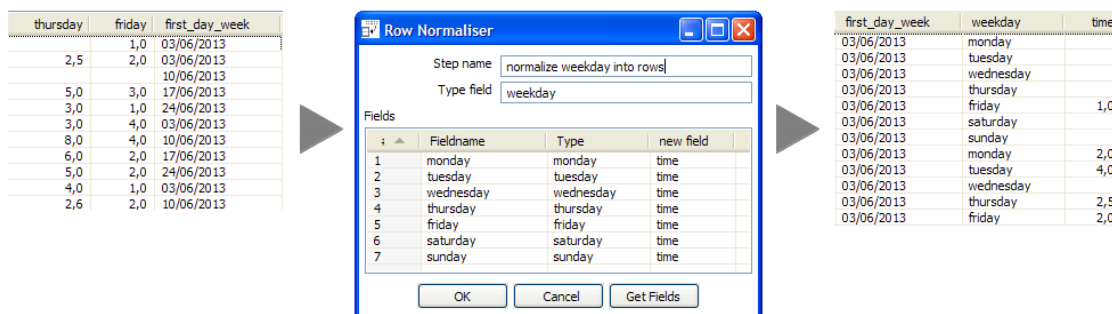
Figura 24 - Processo de ETL no carregamento da tabela de Factos (fonte: própria)



Em termos de tratamento dos dados-fontes, será curial referir três passos que exigiram um tratamento mais específico:

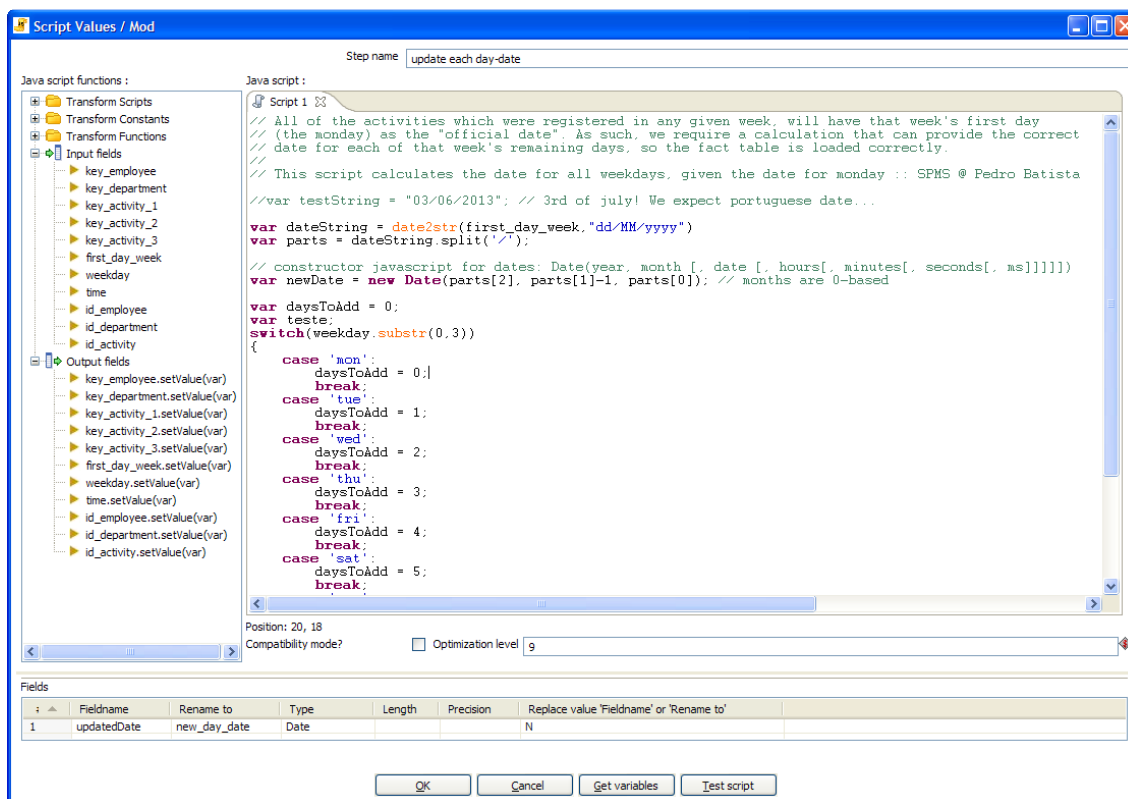
1. A necessidade de normalizar o campo do dia da semana, dado que os dias da semana estão representados nos dados-fonte como colunas e não como linhas de registo, conforme demonstra a próxima figura:

Figura 25 - Normalização do campo "dia", de colunas para linhas (fonte: própria)



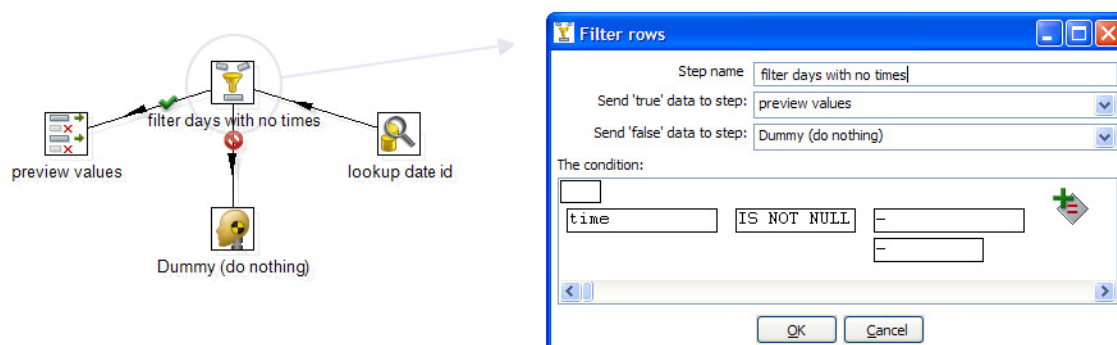
2. O facto de todos os tempos registados no sistema fonte estarem referenciados apenas ao primeiro dia da respetiva semana (data da segunda-feira), obrigou a atualizar as datas dos restantes dias, com recurso a um pequeno procedimento javascript:

Figura 26 - Atualização das datas de cada dia da semana (fonte: própria)



3. Em resultado da normalização dos dias da semana, referida acima, os registos de tempo ficarem com uma densidade ainda mais esparsa do que aquela evidenciada nos dados-fonte, tendo sido necessário colocar um filtro no processo de ETL para evitar carregar a tabela de factos com registos de tempos nulos (vazios na fonte).

Figura 27 - Filtragem de datas sem registos de tempo (fonte: própria)



Os restantes passos do ETL para esta iteração foram todos de execução regular, consistindo essencialmente na obtenção das diversas chaves estrangeiras para cada linha de registo, através da referência às respetivas tabelas de dimensão.

4.1.3. Reporting

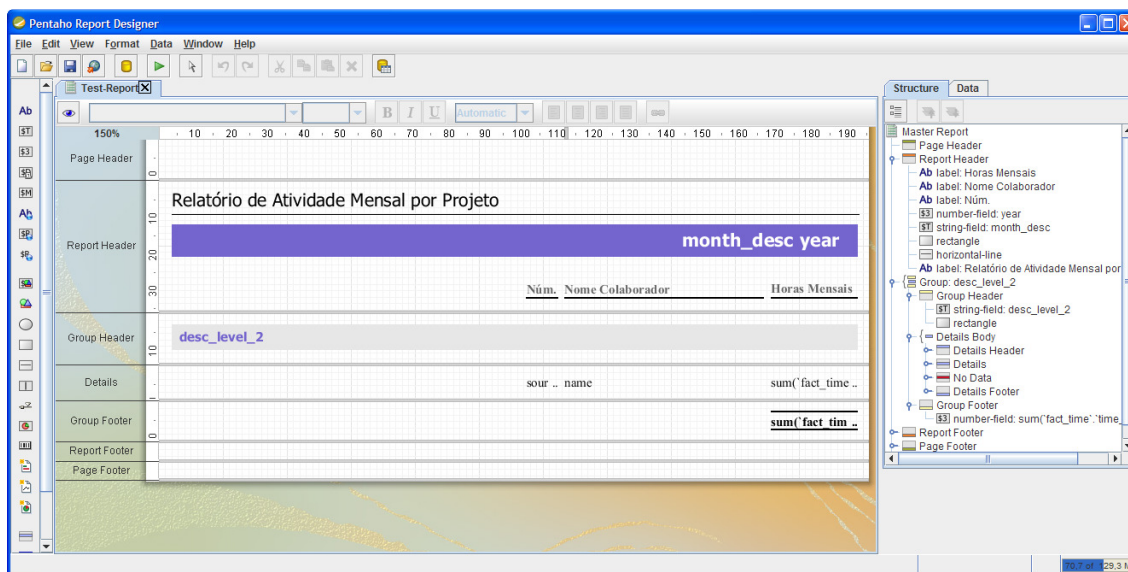
A deliberação prévia com o cliente tinha estabelecido a necessidade mínima de reporting, que permitiria provar que a primeira iteração desta solução conseguiria replicar o único relatório já existente. Foi elaborado, em conjunto, um esboço do que devia ser o primeiro relatório:

Figura 28 - Esboço do relatório a produzir na iteração # 0 (fonte: própria)

Relatório de Atividade Mensal - Por Projeto		
Mês de JUNHO 2013		
Projeto AAA		
101 - Colaborador SG	23,0 hrs	
103 - Colaborador JN	4,5 hrs	
	27,5 hrs	
Projeto BBB		
103 - Colaborador NG	86,0 hrs	
105 - Colaborador TR	12,5 hrs	
122 - Colaborador AE	20,0 hrs	
	118,5 hrs	

A implementação deste relatório, com recurso à ferramenta Pentaho Report Designer (PRD), não apresentou dificuldades relevantes, conforme é possível observar na seguinte figura:

Figura 29 - Elaboração do primeiro relatório na ferramenta Pentaho PRD (fonte: própria)



É importante relevar que a ligação de dados (data source) foi estabelecida diretamente ao data mart. Desta forma, não foram aproveitadas duas funcionalidades relevantes do Pentaho BI: a possibilidade de utilizar fontes de dados definidas no servidor (e não embutidas, com recurso a driver JDBC, no próprio relatório), e a possibilidade de elaborar relatórios por cima de uma camada de metadados. No espírito de desenvolvimento Agile, estas funcionalidades serão exploradas em iterações futuras.

4.1.4. Plataforma de BI (Pentaho BI Server)

Para concluir a primeira iteração do sistema “BI Gestão de Projetos”, faltava ainda publicar o relatório para a plataforma de BI – o Pentaho BI Server, para que pudesse ser acedido sem necessidades de ferramentas de desenvolvimento (ou seja, sem ser através do Report Designer).

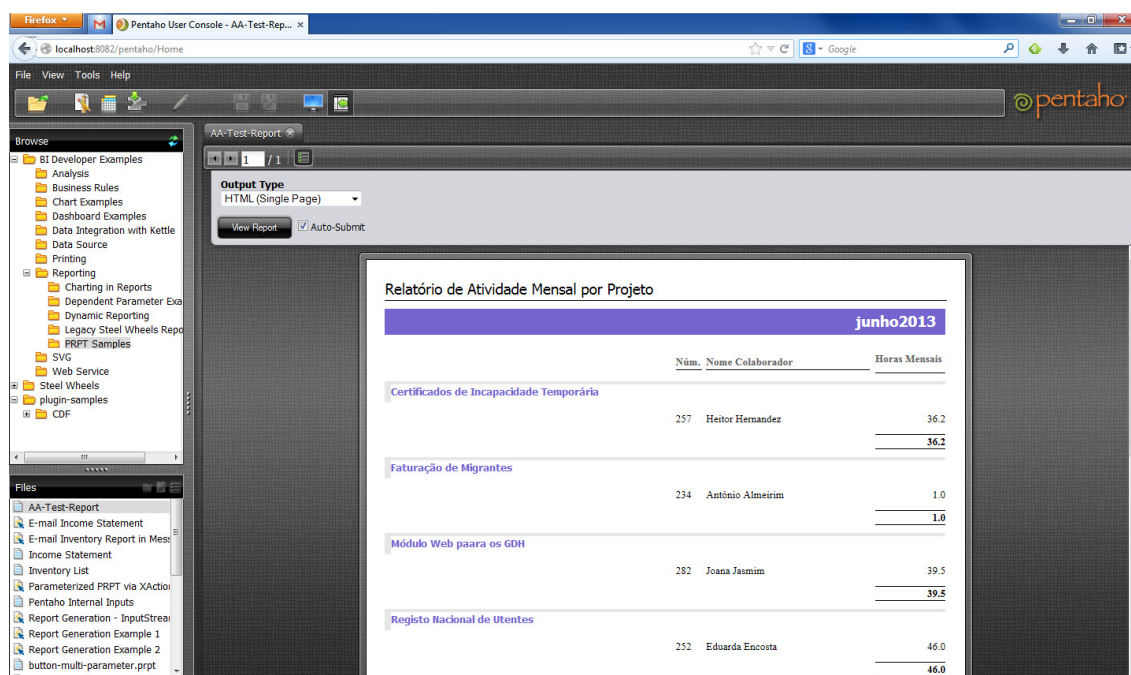
Mais uma vez, foram ignoradas diversas facetas da plataforma BI que, em iterações futuras, deverão ser consideradas, nomeadamente:

- Repositório de soluções – não foi estabelecido ou configurado nenhum repositório de solução específico para esta iteração, sendo aproveitado um dos dois repositórios de demonstração, pré-instalados no BI Server.
- Não foram definidas sequências de ação específicas para este relatório, à parte da ação por defeito: *executar relatório*, não sendo possível passar parâmetros (filtros, por exemplo) para o relatório.

- Criação e gestão de utilizadores – também não foram criados utilizadores específicos para esta iteração, aproveitando-se novamente os utilizadores por defeito na instalação do Pentaho BI Server.

Após a publicação do relatório, executada diretamente do Report Designer, foi possível aceder ao mesmo através do acesso web ao BI Server, conforme exemplificado na Figura 30:

Figura 30 - Acesso ao relatório dentro do Pentaho BI Server (fonte: própria)



The screenshot shows the Pentaho BI Server web interface. On the left is a navigation pane with a tree view of report categories like 'BI Developer Examples', 'Reporting', and 'Files'. The main area displays a report titled 'Relatório de Atividade Mensal por Projeto' for 'junho2013'. The report is in HTML format and contains a table with the following data:

	Núm.	Nome Colaborador	Horas Mensais
Certificados de Incapacidade Temporária			
	257	Heitor Hernandez	36.2
			36.2
Faturação de Migrantes			
	234	António Almeirim	1.0
			1.0
Módulo Web para os GDH			
	282	Joana Jasmin	39.5
			39.5
Registo Nacional de Utentes			
	252	Eduarda Encosta	46.0
			46.0

Por defeito, o BI Server possibilita a extração do relatório em diversos formatos, incluindo pdf, html, excel (xls e xlsx) e texto. Não foi efetuada nenhuma otimização para qualquer destes formatos, sendo que a qualidade de apresentação, tanto em pdf como html, foi considerada bastante satisfatória, para uma primeira iteração.

4.1.5. Resumo da Iteração # 0

Esta primeira iteração foi bem-sucedida, não tendo revelado problemas inesperados. Dado o conhecimento prévio detido sobre algumas das ferramentas utilizadas – ainda que sem grande grau de profundidade técnica – não era exetável encontrar dificuldades, no que se resumiu a uma pequena prova de conceito, com dados artificiais. Ainda assim, serviu para reforçar a opinião de que a plataforma Pentaho servirá os propósitos desejados, em termos de projeto BI-GEST.

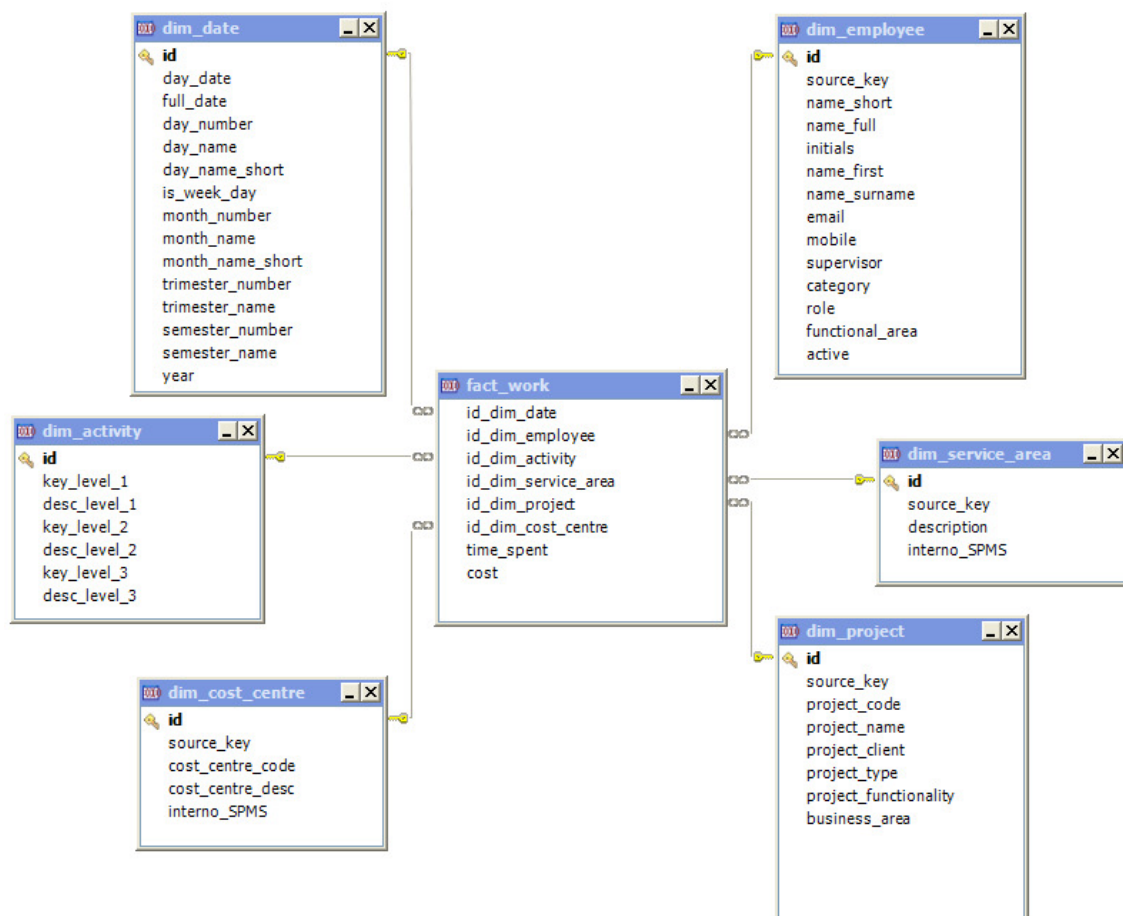
4.2. ITERAÇÃO # 1 – APERFEIÇOAMENTO DO MODELO DE DADOS

A fase seguinte da realização deste projeto incidiu sobre a adaptação do modelo de dados, assegurando que o mesmo cobria todas as necessidades informativas reais, no âmbito da exploração de informação sobre a gestão de projetos na SPMS. Deste modo, as principais tarefas desta iteração concentraram-se no redesenho do data warehouse e na redefinição e aperfeiçoamento dos processos de ETL.

4.2.1. Desenho do Data Warehouse

Em relação ao modelo de dados testado na iteração inicial, decidiu-se manter o esquema em estrela (star schema), sendo apenas necessário incrementar o número de dimensões utilizadas, melhorar as métricas da tabela de factos e adicionar mais detalhe a cada dimensão (através dos respetivos campos de atributos), conforme se observa na seguinte figura:

Figura 31 - Diagrama do data mart para esta iteração (fonte: próprio)



Em termos concretos, foram realizadas as seguintes alterações ao desenho do data mart:

- Revisão à nomenclatura de todos os objetos, em consonância com o alargamento de âmbito introduzido pelo novo modelo de dados; por exemplo, a tabela de factos mudou de “tempo” (fact_time), quando apenas registava as horas realizadas, para “trabalho” (fact_work), agora que regista tanto horas como o custo do trabalho realizado.
- Identificadas duas novas dimensões, sendo que a nova dimensão “projeto” resulta da subdivisão da dimensão “atividade”, que já existia anteriormente, enquanto a dimensão “centro de custo” foi identificada pela primeira vez nesta iteração.
- Adicionada uma nova métrica à tabela de factos, permitindo medir o custo do trabalho inerente a cada registo; desta forma, continuamos a contar somente com factos puramente aditivos, e apresentado a mesma granularidade.
- Dimensões enriquecidas com novos atributos, permitindo melhorar a capacidade de exploração destes dados por parte dos futuros utilizadores; foram alvo de especial atenção as dimensões “colaborador” (dim_employee), atividade (dim_activity) e projeto (dim_project).
- A especificação de atributos adicionais permitiu, adicionalmente, que pelo menos duas dimensões (data e atividade) fossem construídas com uma estrutura hierárquica interna. Deste modo, será possível aproveitar estas hierarquias em fases posteriores, tanto ao nível de reporting como em análises multidimensionais.
- Dado que a partir desta fase, lidava-se com dados reais, foi sendo necessário auditar e inspecionar a gama de valores-limite para cada um dos atributos dimensionais, por forma a poder modelar a respetiva representação física em sistema de base de dados, sem correr riscos de violações de integridade e/ou tipo de dados.

4.2.2. Processos de ETL

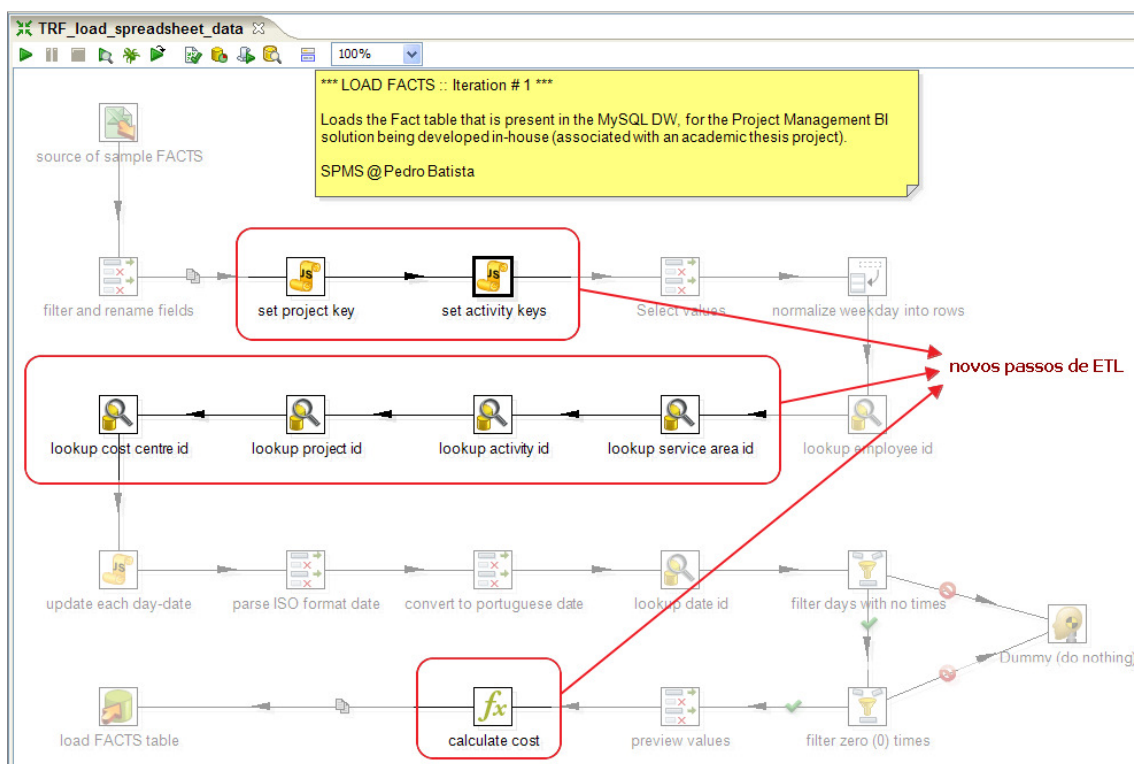
Praticamente todos os processos de ETL, desenvolvidos durante a iteração inaugural, puderam ser reaproveitados. No entanto, o facto de estarmos agora a lidar com o universo completo de dados reais trouxe dificuldades acrescidas à extração, tratamento e carregamento dos dados-fonte. Essencialmente, detetaram-se diversos problemas relacionados com a qualidade dos dados-fonte, nomeadamente em termos de valores incongruentes face às respetivas dimensões e dados duplicados, pelo que mesmos os passos de ETL existentes tiveram que ser ajustados.

Para além de ajustamentos nos passos existentes, tiveram que ser desenvolvidos alguns novos passos, em consequência de ser necessário:

1. Carregar diretamente as dimensões, a partir de listagens mestre;
2. Ajustar o carregamento da tabela de factos, para:
 - a. Incorporar a associação às novas dimensões.
 - b. Permitir separar “projetos” das “atividades”, carregando ambos os registos de forma independente.
 - c. Apurar e carregar a nova métrica de custos.

No processo de carregamento da tabela de factos, por exemplo, foi necessário desenvolver 7 novos passos, conforme revela a Figura 32:

Figura 32 - Processo de ETL melhorado, para carregamento de factos (fonte: própria)



Nesta fase, e em resultado de reflexão conjunta com o cliente deste projeto, foi decidido não incorporar, por enquanto, *slowly-changing dimensions* no atual modelo. Todavia, ficou claro que algumas das dimensões em uso necessitarão, obrigatoriamente, de refletir histórico de alterações, pelo que esta gestão terá que ser reequacionada numa futura iteração.

4.2.3. Reporting

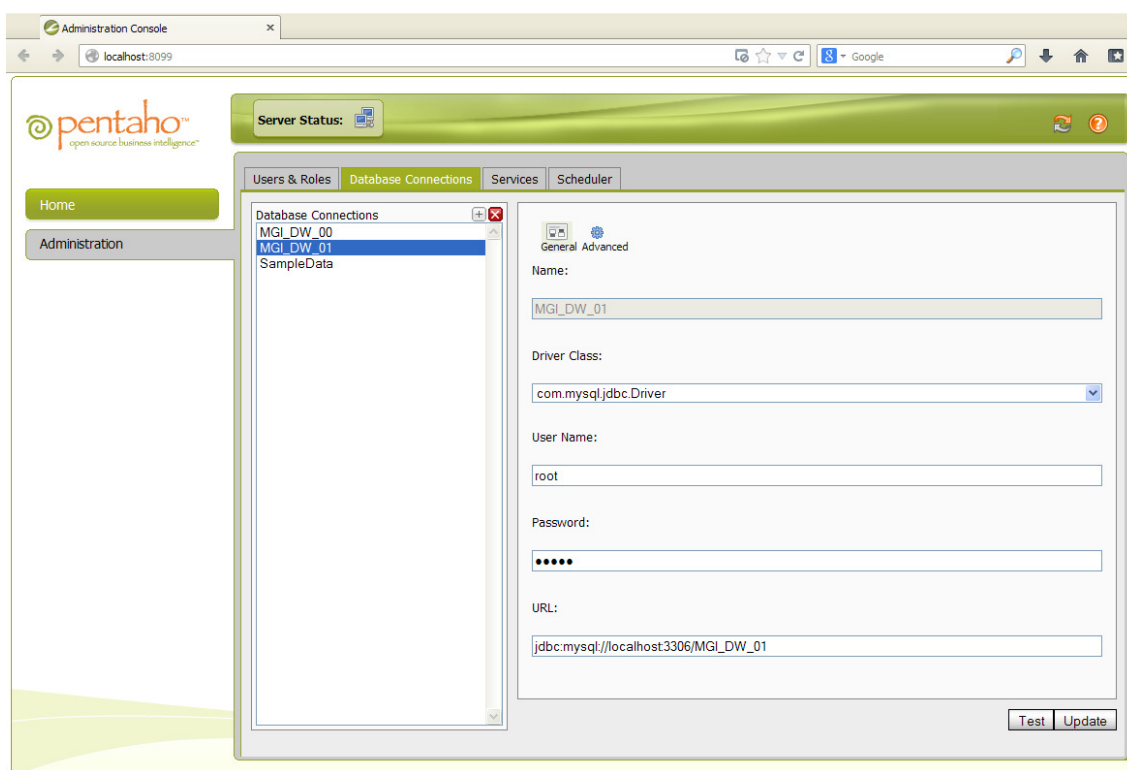
Nesta iteração, não foram desenvolvidos relatórios através do Pentaho Report Designer. A premissa base deste segundo ciclo de desenvolvimento era conseguir

dotar o utilizador de capacidade para produzir autonomamente os seus relatórios (o chamado *self-service reporting*). Nesse sentido, o esforço de reporting, nesta iteração, centrou-se na configuração das funcionalidades nativas da plataforma de BI da Pentaho, conforme descrevemos de seguida.

4.2.4. Plataforma de BI (Pentaho BI Server)

O primeiro passo para estabelecer a funcionalidade de reporting inerente à plataforma de BI da Pentaho, é a configuração de uma conexão à base de dados do data mart. Esta configuração realiza-se por via do Pentaho Administration Console (PAC), que constitui um componente separado da restante plataforma, dedicado exclusivamente à administração de acessos, conexões e algumas outras funcionalidade administrativas, conforme apresentamos na figura seguinte:

Figura 33 - Configuração de conexão a base de dados no Administration Console (fonte: própria)

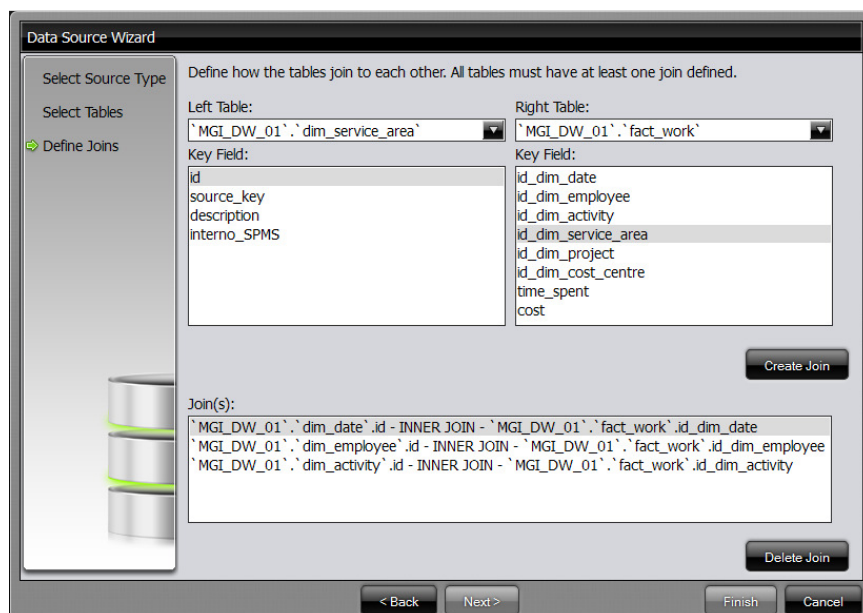


Note-se que esta conexão ficará configurada e alojada no próprio servidor de BI, podendo ser partilhada mais tarde, para efeitos de reporting e análises multidimensionais (cubos Mondrian).

Após ter-se configurado uma conexão à base de dados do data mart, torna-se possível configurar uma fonte de dados (data source) no Pentaho User Console, que será utilizado pelos próprios utilizadores quando da construção dos seus relatórios. A configuração desta fonte de dados é conseguida através de um wizard de 3 passos,

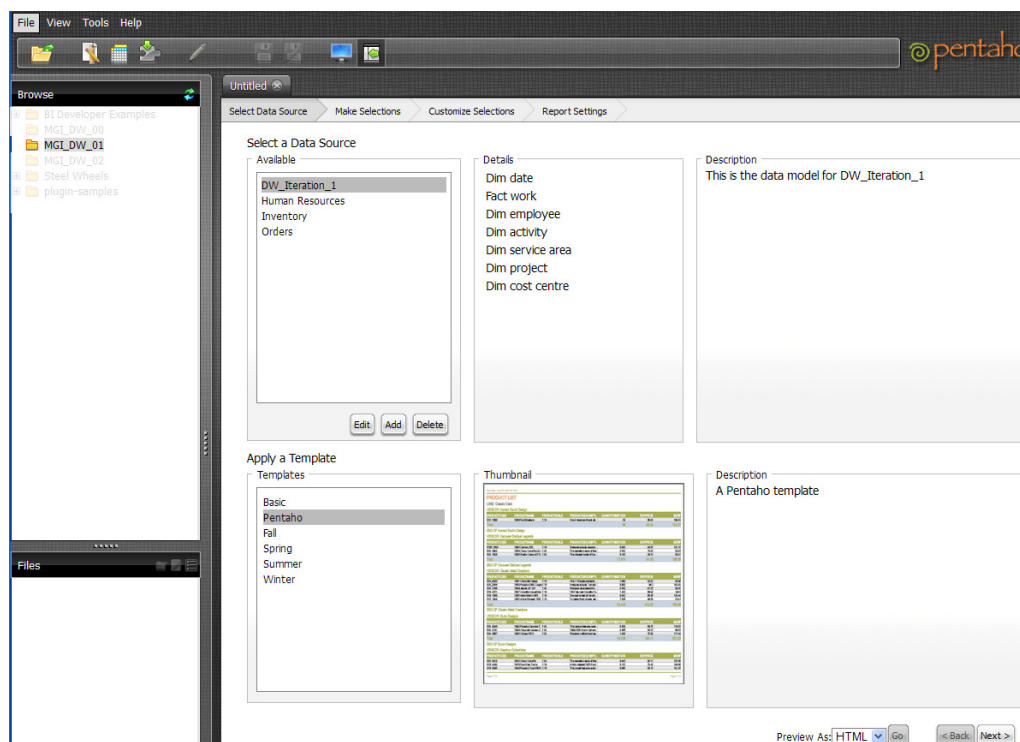
sendo necessário discriminar quais as tabelas que serão utilizados para efeitos de reporting e como estão unidas (joined), conforme é possível observar no terceiro e último ecran do wizard:

Figura 34 - Criação da fonte de dados "DW_Iteration_1" para reporting (fonte: própria)



Após estabelecer uma fonte de dados (neste caso, como "DW_Iteration_1"), a seleção de "Novo Relatório" irá abrir novo wizard, permitindo construir relatórios na ferramenta Web Ad Hoc Query and Reporting (WAQR), conforme mostra a Figura 35:

Figura 35 - Construção de novos relatórios no WAQR (fonte: própria)

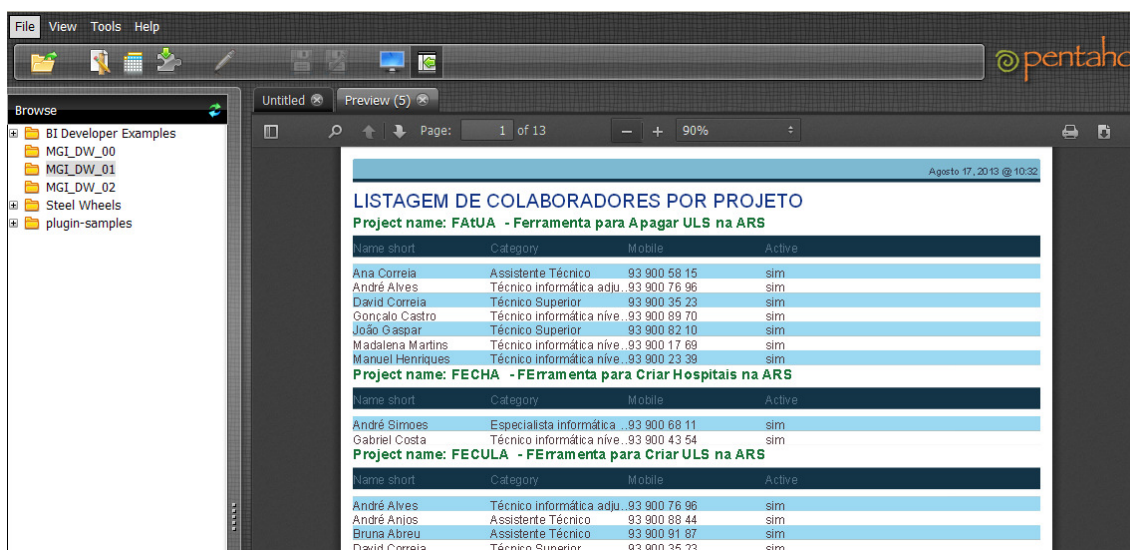


Os 4 passos de interação deste wizard permitem ao utilizador realizar as seguintes ações:

1. **Escolha de fonte de dados:** neste ecrã, é possível selecionar a fonte de dados e a aplicação de um de 6 modelos de formatação pré-definidos;
2. **Selecionar campos:** permite escolher quais os campos de informação que devem compor o relatório, e de que forma devem ser agrupados ou filtrados;
3. **Personalização do relatório:** permite uma ligeira configuração de cada um dos campos selecionados previamente (essencialmente, o formato se for numérico, e a ordenação dos resultados);
4. **Configuração do relatório:** permite escolher o tamanho da página associada ao relatório e inserir texto estático para as zonas de cabeçalho e rodapé;

Após concluir os passos no wizard, cabe ao utilizador escolher em qual dos formatos pretende produzir o relatório: html, pdf, excel ou csv. A figura em baixo evidencia uma simples listagem de colaboradores, obtida através do WAQR:

Figura 36 - Obtenção de relatório WAQR com listagem de colaboradores (fonte: própria)



Agosto 17, 2013 @ 10:32

LISTAGEM DE COLABORADORES POR PROJETO

Project name: FATUA - Ferramenta para Apagar ULS na ARS

Name short	Category	Mobile	Active
Ana Correia	Assistente Técnico	93 900 58 15	sim
André Alves	Técnico informática adju.	93 900 76 96	sim
David Correia	Técnico Superior	93 900 35 23	sim
Gonçalo Castro	Técnico informática níve.	93 900 89 70	sim
João Gaspar	Técnico Superior	93 900 82 10	sim
Madalena Martins	Técnico informática níve.	93 900 17 69	sim
Manuel Henriques	Técnico informática níve.	93 900 23 39	sim

Project name: FECHA - Ferramenta para Criar Hospitais na ARS

Name short	Category	Mobile	Active
André Simoes	Especialista informática	93 900 68 11	sim
Gabriel Costa	Técnico informática níve.	93 900 43 54	sim

Project name: FECULA - Ferramenta para Criar ULS na ARS

Name short	Category	Mobile	Active
André Alves	Técnico informática adju.	93 900 76 96	sim
André Anjos	Assistente Técnico	93 900 88 44	sim
Bruna Abreu	Assistente Técnico	93 900 91 87	sim
David Correia	Técnico Superior	93 900 35 23	sim

4.2.5. Análise de Resultados - Iteração # 1

Esta segunda iteração permitiu trazer alguma robustez ao desenho da solução de data mart para este projeto, acrescentando-lhe o conjunto mínimo de dimensões para que almeje a colmatar necessidades reais da organização. Para além do redesenho do data mart, a provisão de relatórios estruturados à solução permite satisfazer muitas das solicitações iniciais do cliente, permitindo acrescentar valor de uma forma célere e relativamente fácil.

4.3. ITERAÇÃO # 2 – IMPLEMENTAÇÃO DE CUBO MULTIDIMENSIONAL

Nesta fase do projeto, o cliente sentia necessidade de conseguir explorar os valores de tempo e custo contidos na solução já desenvolvida. No fundo, tratava-se de obter capacidade de análise ad hoc sobre estes dados, ou seja, ter acesso a funcionalidades de Online Analytical Processing (OLAP). Nesse sentido, o terceiro ciclo de desenvolvimento centrou-se em dotar o projeto de capacidades OLAP, com recurso ao Pentaho Analysis Services (PAS).

4.3.1. Pentaho Analysis Services (PAS)

O Pentaho Analysis Services é composto por 4 componentes distintos, conforme discriminado na seguinte tabela:

Tabela 5 - Discriminação dos principais componentes do Pentaho Analysis Services (fonte: própria)

Nome	Função	Utilizado no projeto?
JPivot	Camada de visualização e exploração dos cubos multidimensionais, embutida no Pentaho User Console	Sim
Motor ROLAP Mondrian	Motor de OLAP Relacional, permitindo providenciar conjuntos multidimensionais às análises definidas através do JPivot	Sim
Schema Workbench	Ferramenta que permite definir, através de interface gráfico, schemas para cubos Mondrian	Sim
Aggregate Designer	Ferramenta que permite definir, através de interface gráfico, tabelas agregadoras que melhorem o desempenho do cubo	Não

Note-se que o volume dos dados deste projeto não exige, por enquanto, tabelas com dados pré-agregados, pelo que não se equacionou a utilização do Aggregate Designer.

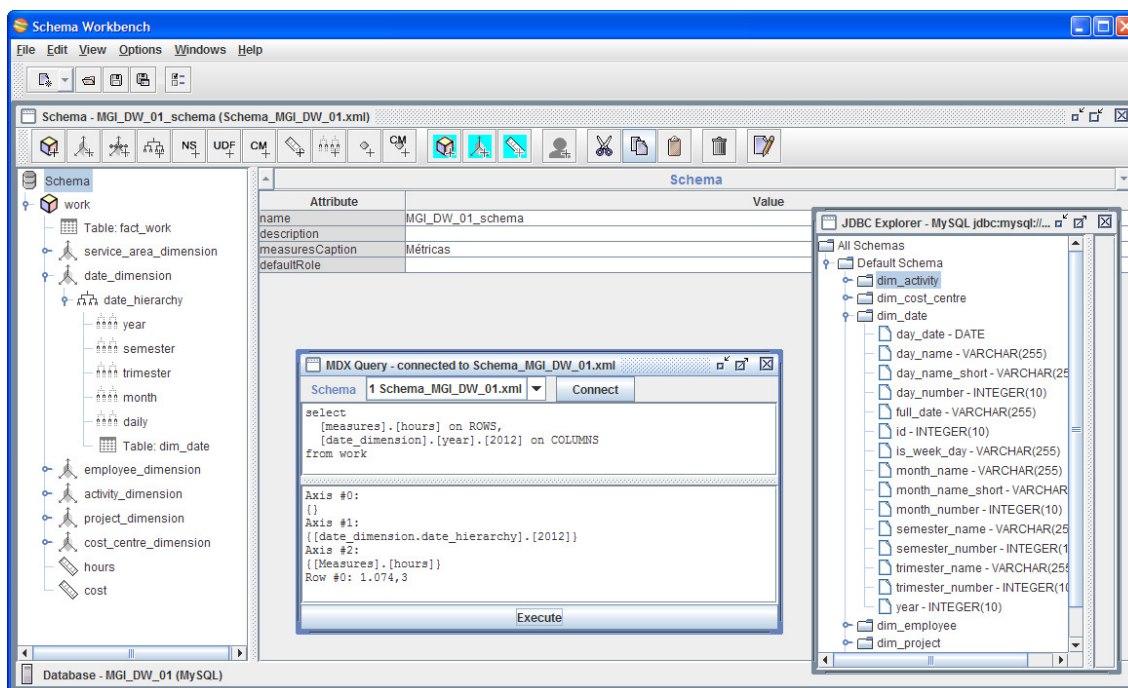
A instalação do Pentaho BI server já inclui, por defeito, os dois primeiros componentes (JPivot e motor ROLAP Mondrian), sendo as duas últimas ferramentas apenas formas mais amigáveis de configuração do cubo OLAP. Na verdade, bastaria definir um schema em linguagem XML e inclui-lo no respetivo diretório da solução para que este fique disponível ao utilizador. Todavia, esta tarefa é bastante facilitada pela ferramenta Schema Workbench, razão pela qual foi utilizada neste projeto.

4.3.2. Definição de schema do cubo multidimensional

O Schema Workbench é um utilitário de uso stand-alone, com capacidades básicas para criação, edição e revisão de schemas. Para além da funcionalidade principal de desenho de schemas, disponibiliza um editor muito elementar para testar expressões MDX e um explorador (também bastante rudimentar) das tabelas e

campos da base de dados associada ao respetivo schema, conforme apresentamos na Figura 37:

Figura 37 - Janelas principais da ferramenta Pentaho Schema Workbench (fonte: própria)



A criação do schema correspondente a este projeto baseou-se em alguns pressupostos, decididos em função das necessidades do cliente e em conjugação com a prerrogativa de manter o desenvolvimento do projeto ágil, flexível e rápido. Deste modo, apenas se definiu um único cubo, com as características que se apresentam na seguinte tabela:

Tabela 6 - Características do cubo multidimensional do projeto (fonte: própria)

Eixo	Tipo	Níveis Hierárquicos	Agregação
Data	dimensão	Ano Semestre Trimestre Mês Dia	--
Colaborador	dimensão	Colaborador	--
Área de Negócio	dimensão	Área de Negócio	--
Projeto	dimensão	Projeto	--
Atividade	dimensão	Atividade Nível 1 Atividade Nível 2 Atividade Nível 3	--
Centro de Custo	dimensão		--
Horas Realizadas	métricas		soma
Custo			soma

Foram integradas no desenho do cubo todas as dimensões constantes do data mart, e configuradas duas hierarquias: a hierarquia temporal clássica (aqui com a

Destaque-se a barra de ferramentas, sendo que abordaremos apenas algumas das principais funcionalidades, tidas como relevantes neste projeto:

Figura 40 - Barra de ferramentas do Pentaho JPivot (fonte: própria)



1. Funcionalidades dedicadas à configuração do cubo:
 - a. Permite trocar os membros do cubo entre os dois eixos, ainda como colocar um ou vários membros como filtro ("slicing the cube");
 - b. Permite obter e interagir diretamente sobre a query MDX que está na origem da matriz apresentada pelo JPivot;
 - c. Configuração da tabela OLAP, ao nível de opções de ordenação;
2. Funcionalidades dedicadas à apresentação visual da matriz OLAP.
3. Opções de drill, sendo as primeiras 3 apenas aplicáveis às dimensões, e a última apenas aplicável às métricas:
 - a. Drill nos membros, em que todas as instâncias desse membro (representadas na matriz) são expandidas;
 - b. Drill na posição, em que apenas a instância selecionada do membro é expandida, ficando as restantes instâncias compactadas;
 - c. Drill de substituição, em que a instância escolhida para drill é substituída pelas respetivas instâncias subalternas, nessa mesma posição;
 - d. Drill pelo centro (aos resultados), sendo apresentados todos os registos da tabela de factos, que contribuíram para os valores constantes da seleção escolhida. Veja-se o seguinte exemplo:

Figura 41 - Representação da funcionalidade de "drill through" (fonte: própria)

Drill Through Table for time_spent									
description	id	year	year	semester_name	semester_number	trimester_name	trimester_number	month_name	month_number
Produção Suporte	20	2012	2012	2º Semestre	2	4º Trimestre	4	dezembro	12
Produção Suporte	20	2012	2012	2º Semestre	2	4º Trimestre	4	dezembro	12

id	name_short	id	desc_level_1	key_level_1
186	Call Center	1	Call Center	Call Center
145	Call Center	1	Call Center	Call Center

Page 1/1 Rows/page 10

4. Opção para obter uma representação em gráfico, dos resultados constantes da tabela OLAP.
5. Opções sobre exportação e impressão da matriz de resultados OLAP.

4.3.4. Funcionalidades OLAP não utilizadas

Será pertinente referir que, tendo em conta a natureza iterativa deste projeto, nem todas as funcionalidades OLAP disponíveis na plataforma Pentaho / Mondrian foram exploradas. No entanto, há algumas que aparentam revelar bastante potencial, e poderão ser incorporadas em futuras iterações do projeto, nomeadamente:

- Controlos de acesso por via de configuração de ‘roles’, permitindo definir “quem” tem acesso a “que” conjuntos de dados;
- Formatação condicional, através da inclusão de fórmulas nas ‘strings’ de formatação;
- Membros calculados;

4.3.5. Análise de Resultados - Iteração # 2

A capacidade de poder explorar e extrair informação, de forma autónoma e sem necessidade de intervenção por parte da equipa técnica, é algo que os utilizadores do sistema BI-GEST valorizam muito. Todavia, e muito embora esta iteração tenha tido sucesso do ponto de vista estritamente técnico – afinal, é agora possível construir e executar operações OLAP clássicas (slice, dice, drill) sobre os custos e horas dos projetos SPMS – esta iteração não se traduziu num sucesso junto dos utilizadores.

A razão principal pelo qual a adição de capacidade OLAP não recebeu a aceitação esperada prende-se, essencialmente, com a dificuldade em usar o interface JPivot para efeitos de análises OLAP. Os utilizadores manifestaram desagrado e frustração com a elevada curva de aprendizagem do JPivot, não achando que esta interface seja aceitável do ponto de vista de usabilidade para quem não seja utilizador avançado (com noções de OLAP e multidimensionalidade).

4.4. ITERAÇÃO # 3 – UTILIZAÇÃO DE FERRAMENTAS MELHORADAS

Conforme referido, após ter alcançado esta fase de desenvolvimento, a preocupação principal dos utilizadores e clientes deste sistema passava pela fraca usabilidade das ferramentas nativas de exploração de dados – tanto de reporting (WAQR) como de análise (JPivot). Efetivamente, as funcionalidades nativas de exploração da plataforma Pentaho, na sua versão comunidade, não são particularmente apelativas – não tanto por não serem eficazes, mas pela baixo grau de usabilidade e elevada curva de aprendizagem (especificamente no que concerne o JPivot). Aliás, será relevante recordar que a oferta comercial da Pentaho oferece ferramentas alternativas (e muito melhoradas) de exploração de dados, como é o caso do Pentaho Analyzer (apenas disponível na versão comercial).

A comunidade existente em redor deste ecossistema de business intelligence open source também teve consciência desta lacuna, e como forma de colmatar esta necessidade, têm surgido diversas iniciativas open source para criar ferramentas melhoradas de exploração de dados na plataforma Pentaho.

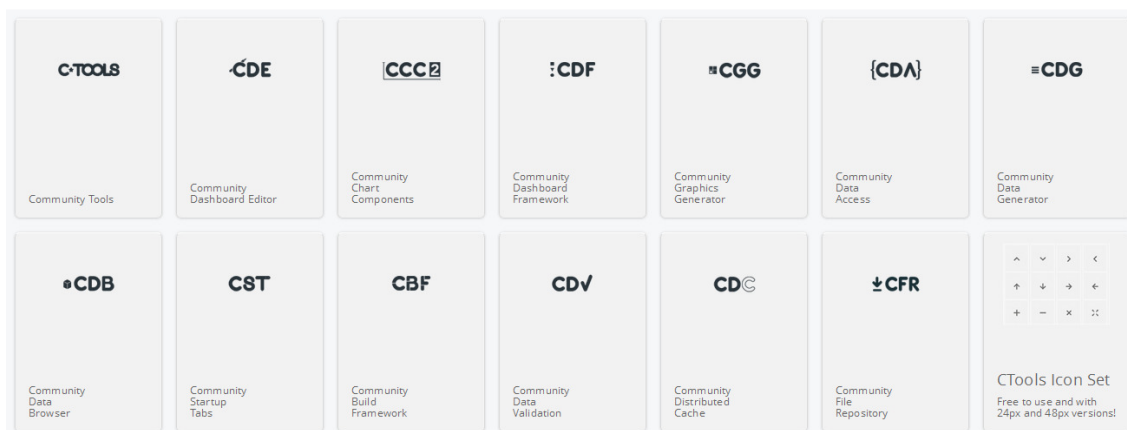
Assim, colocamos como meta para este novo ciclo de desenvolvimento – no que será a iteração # 3 – disponibilizar aos utilizadores uma solução melhorada, recorrendo à integração de novas ferramentas open source para a exploração de dados.

4.4.1. Apresentação das Ferramentas CTools

Um dos pacotes de ferramentas integráveis (add-on) mais conhecidos para o Pentaho BI é a suite Community Tools (CTools). Este pacote de ferramentas nasce do trabalho de uma empresa portuguesa, que em agosto de 2008 desenvolve um componente para facilitar a criação e usabilidade de dashboards no Pentaho BI. Conforme relatado num infográfico da história dos componentes CTools (Webdetails), a força motivadora para a criação do Community Dashboard Framework (CDF) prendeu-se com o desejo de incrementar o impacto visual dos dashboards, que estaria a prejudicar a competitividade da plataforma community da Pentaho. Um ano depois, em setembro de 2009, disponibilizam novo componente, o Community Dashboard Editor, facilitando ainda mais a criação e edição de dashboards assentes sobre a framework CDF.

Após a adoção como funcionalidade nativa da plataforma Pentaho na sua versão 3.0, a empresa Webdetails continuou a alargar rapidamente a sua suite de componentes open source, lançando add-ons para várias facetas da plataforma: gráficos, conexões de dados, compilação, etc.

Figura 42 - Família de componentes na suite CTools (fonte: Webdetails)



Dos vários componentes disponibilizados na suite CTools, talvez os mais emblemáticos, e que maior utilização recebem, sejam os seguintes:

- i. **Community Dashboard Framework (CDF)** – até à criação desta framework, o desenvolvimento de dashboards na plataforma Pentaho era laborioso, complexo e exigia conhecimentos relativamente profundos de tecnologias web (http, html, css e javascript). Esta ferramenta veio unir todas estas camadas num só modelo, abstraindo toda essa complexidade através de um framework moderno, extensível e baseado em normas abertas.
- ii. **Community Dashboard Editor (CDE)** – muito embora a framework CDF tenha facilitado a criação programática de dashboards, não era verdadeiramente um editor gráfico de dashboards. Para colmatar essa falta, foi criado o CDE, que se assume como um editor gráfico integrado para criar, editar e pré-visualizar dashboards Pentaho.
- iii. **Community Data Access (CDA)** – esta ferramenta foi desenvolvida como forma de oferecer uma camada de abstração sobre as diversas fontes de dados tipicamente utilizadas em soluções de BI na plataforma Pentaho. Assim, torna-se possível cruzar e misturar dados oriundos de fontes tão diversas como bases de dados (SQL), cubos (MDX) ou mesmo ficheiros (csv, xls, etc.), ao mesmo tempo que se reforçam aspetos essenciais, como segurança, desempenho (cache) e flexibilidade.

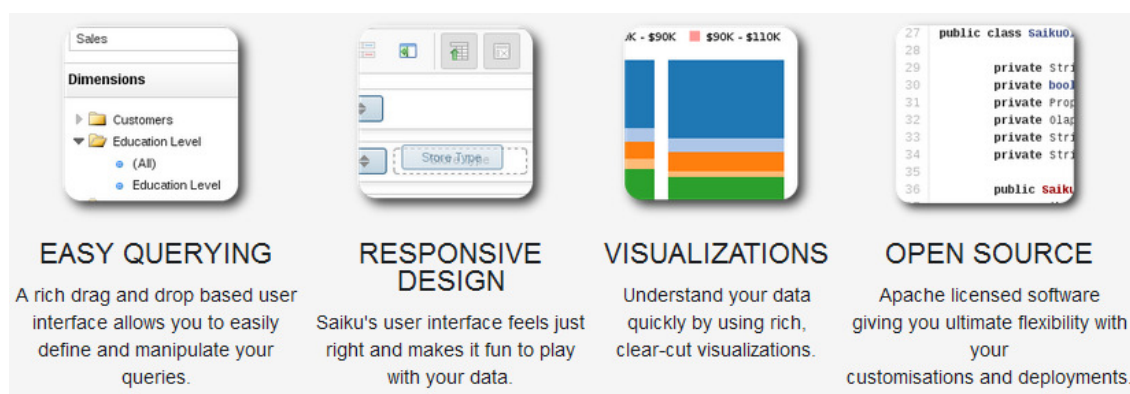
4.4.2. Apresentação das Ferramentas Saiku

Em 2008, foi criada uma ferramenta, inicialmente designada de Pentaho Analysis Tool, para funcionar como uma forma mais amigável de aceder e explorar as capacidades OLAP de cubos multidimensionais suportados pela biblioteca OLAP4J (tipicamente, o motor Mondrian embutido na plataforma Pentaho, mas também

Microsoft SSAS, SAP BW, Oracle Hyperion, entre outros). Tem recebido evoluções através do esforço da comunidade open source e, em 2010 e após uma reestruturação integral, passa a designar-se Saiku.

Essencialmente, o add-on Saiku disponibiliza uma ferramenta analítica, baseada em tecnologia web, que oferece aos utilizadores a capacidade de analisar e criar relatórios, de forma fácil e rápida, com base na informação contida em repositórios multidimensionais.

Figura 43 - Principais funcionalidades da ferramenta Saiku (fonte: Saiku Analytics)

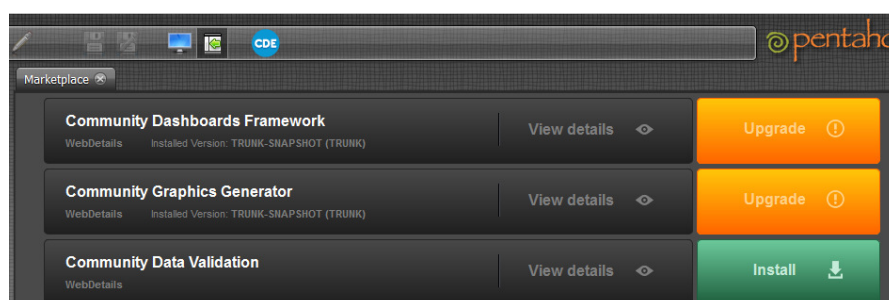


Para além da componente dedicada ao OLAP, surgiu em 2012 um projeto relacionado mas independente, intitulado Saiku Reporting, que oferece uma forma mais desenvolvida e amigável de construir relatórios ad-hoc pelos próprios utilizadores, assente na camada de metadata (se existir) do Pentaho. Desta forma, é possível melhorar as capacidades da ferramenta nativa de reporting desta plataforma (o componente WAQR, visto na Iteração # 1).

4.4.3. Instalação das Ferramentas

A instalação destas ferramentas tem melhorado significativamente, já não sendo necessário – embora ainda sejam opções válidas – a utilização de scripts ou de passos manuais de inserção/substituição de ficheiros. A própria plataforma Pentaho inclui a funcionalidade de Marketplace, permitindo escolher que add-ons instalar ou atualizar:

Figura 44 - Utilização do Marketplace para instalação de componentes (fonte: própria)



4.4.4. As Novas Ferramentas: Revisitando Reporting e Análise

Podemos ver, em baixo, as principais diferenças do Saiku Reporting face à ferramenta nativa do Pentaho, para o mesmo tipo de listagem operacional simples (lista de colaboradores por projeto) apresentada anteriormente na Iteração # 1:

Figura 45 – Relatório com listagem de colaboradores através do Saiku Reporting (fonte: própria)

The screenshot shows the Saiku Reporting application interface. On the left, there is a 'Models' sidebar with a tree view showing dimensions like 'Dim employee', 'Dim activity', 'Dim service area', 'Dim project', and 'Dim cost centre'. The main area displays a report titled 'Listagem de Colaboradores por Projeto'. The report is structured with three sections, each corresponding to a project name: 'FATUA - Ferramenta para Apagar ULS na ARS', 'FECHA - Ferramenta para Criar Hospitais na ARS', and 'FECULA - Ferramenta para Criar ULS na ARS'. Each section contains a table with the following columns: 'Name short', 'Category', 'Mobile', and 'Active'. The data is filtered by 'Project code'.

Name short	Category	Mobile	Active
Ana Correia	Assistente Técnico	93 900 58 15	sim
André Alves	Técnico informática adjunto nível 1	93 900 76 96	sim
David Correia	Técnico Superior	93 900 35 23	sim
Gonçalo Castro	Técnico informática nível 1-grau 1	93 900 89 70	sim
João Gaspar	Técnico Superior	93 900 82 10	sim
Madalena Martins	Técnico informática nível 2-grau 3	93 900 17 69	sim
Manuel Henriques	Técnico informática nível 1-grau 2	93 900 23 39	sim

Name short	Category	Mobile	Active
André Simoes	Especialista informática nível 2-grau 3	93 900 68 11	sim
Gabriel Costa	Técnico informática nível 2-grau 3	93 900 43 54	sim

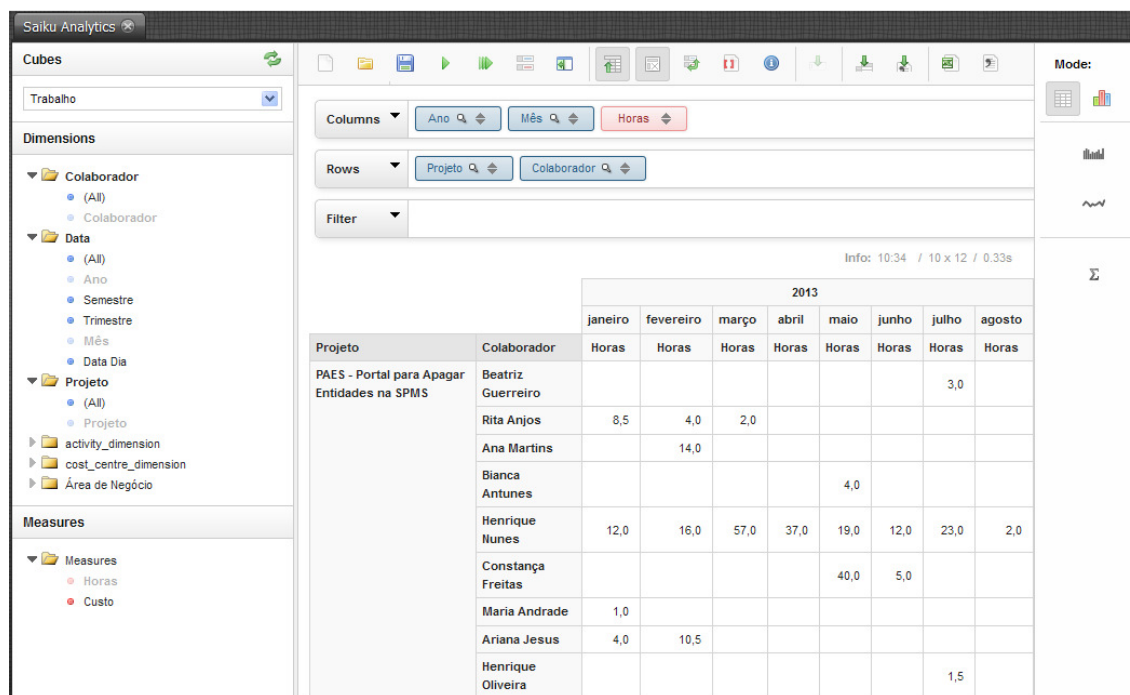
Name short	Category	Mobile	Active
André Alves	Técnico informática adjunto nível 1	93 900 76 96	sim
André Anjos	Assistente Técnico	93 900 18 44	sim

As principais vantagens do Saiku Reporting sobre a ferramenta nativa são:

- A adoção de uma filosofia de “arrastar-e-largar” (drag-and-drop) ao invés da “construção guiada” (wizard), privilegiando a interatividade e experimentação, dado que os resultados ficam imediatamente visíveis;
- Maior flexibilidade em termos de personalização (formatação de cores, fontes, tamanhos, etc.) de elementos individuais do relatório, bem como de seções inteiras;
- A possibilidade de incluir colunas calculadas;
- Permite limitar o número de linhas apresentadas no relatório;
- Maior flexibilidade nos formatos que podem ser exportados, adicionando formatos para o Pentaho Report Designer (.prpt) e CDA (.cda);

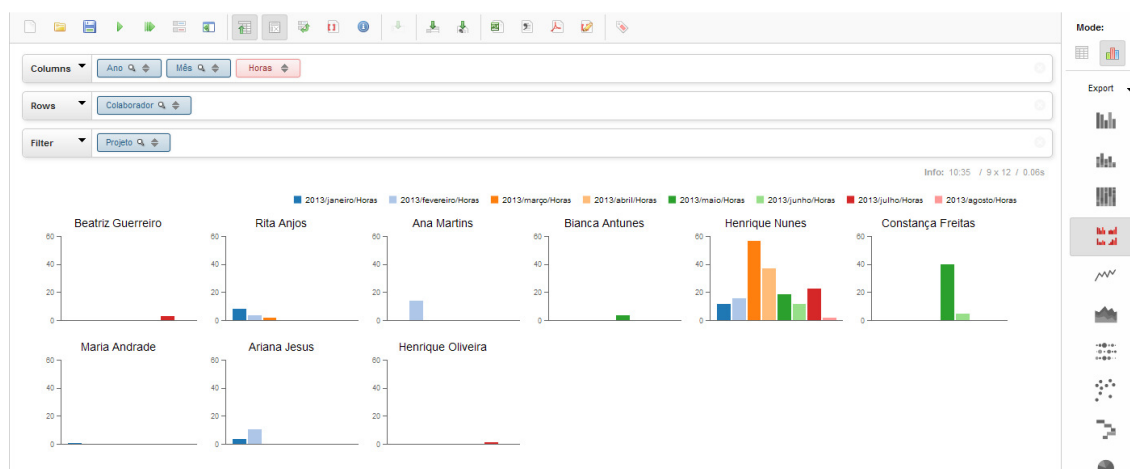
No que concerne a utilização do Saiku como ferramenta de manipulação e exploração de capacidades OLAP, podemos ver em baixo as principais diferenças face ao JPivot (ferramenta nativa na versão comunidade):

Figura 46 - Obtenção de matriz cruzada com o Saiku Analytics (fonte: própria)



Uma das várias funcionalidades que esta ferramenta oferece é a possibilidade de obter, de forma fácil e imediata, uma perspetiva gráfica dos dados apresentados, conforme se demonstra na Figura 47:

Figura 47 - Exploração gráfica dos dados no Saiku Analytics (fonte: própria)



Note-se que, sendo esta ferramenta uma forma alternativa de providenciar as capacidades OLAP associadas ao motor multidimensional Mondrian que é instalado

por defeito na plataforma Pentaho, todas as principais funcionalidades oferecidas pela ferramenta nativa – o JPivot, descrito na Iteração # 2 – estão também garantidas. Assim sendo, não iremos detalhar exaustivamente todas as diferenças entre as duas ferramentas de exploração OLAP, bastando referir as principais vantagens do Saiku:

- Grau de usabilidade muito melhorado face ao JPivot, com uma curva de aprendizagem substancialmente inferior;
- Apresentação melhorada, em resultado de utilização de um interface moderno, assente em normas e padrões web atuais (ao contrário do JPivot, que depende de JavaServer Pages e não é atualizado desde 2008);
- Provisionamento de exploração gráfica significativamente mais flexível e amigável do que a oferecida no JPivot, tornando praticável este tipo de análises para o utilizador comum;
- Maior flexibilidade nos formatos que podem ser exportados, adicionando formato de dados simples (csv) e melhorando a exportação dos já existentes (pdf e xls);
- Acrescenta a funcionalidade obter algumas estatísticas básicas sobre o conjunto de dados que a pesquisa efetuada devolveu do cubo:

Figura 48 - Exemplo de estatísticas básicas no Saiku Analytics (fonte: própria)

Statistics	janeiro / Horas	fevereiro / Horas	março / Horas	abril / Horas	maio / Horas	junho / Horas	julho / Horas	agosto / Horas
Min	1.000	4.000	2.000	37.000	4.000	5.000	1.500	2.000
Max	12.000	16.000	57.000	37.000	40.000	12.000	23.000	2.000
Sum	25.500	44.500	59.000	37.000	63.000	17.000	27.500	2.000
Average	6.375	11.125	29.500	37.000	21.000	8.500	9.167	2.000
Std. Deviation	4.204	4.560	27.500	0.000	14.765	3.500	9.801	0.000

4.4.5. Análise de Resultados - Iteração # 3

Esta iteração teve uma aceitação muito positiva junto dos utilizadores, dado os elevados fatores de usabilidade, flexibilidade e utilização agradável dos componentes add-on que estão agora disponíveis nesta solução.

De facto, será importante salientar que as capacidades OLAP, desenvolvidas durante a iteração # 2, apenas ficam aptas para uma utilização real e frequente, por utilizadores não-especialistas, caso estes componentes melhorados (essencialmente, Saiku Analytics) sejam instalados. A ferramenta JPivot não pode ser considerada como de utilização aceitável para quem não seja experiente em conceitos de OLAP e de estruturas multidimensionais. Deste modo, esta iteração acrescentou um elevado grau de autonomia aos utilizadores desta solução.

4.5. ITERAÇÕES FUTURAS – FUNCIONALIDADES A DESENVOLVER

Conforme referido anteriormente, um projeto desenvolvido sob metodologias Ágeis nunca está verdadeiramente acabado, pois por definição é sempre possível continuar a melhorá-lo iterativamente. Todavia, e sendo esse o caso com a implementação real do BI-GEST na SPMS, para efeitos de estudo académico, temos que o dar por encerrado em algum ponto. Ainda assim, será oportuno referir algumas das possíveis iterações futuras deste projeto, possibilitando o reforço e melhoria das suas funcionalidades.

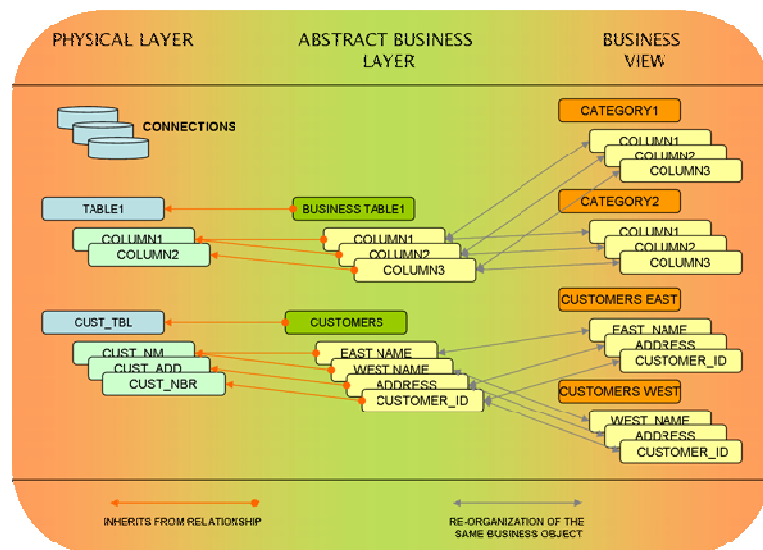
Note-se que algumas destas futuras iterações forem, de facto, já iniciadas em termos do desenvolvimento contínuo deste projeto na SPMS, apenas não sendo possível incorporar esses passos no relatório aqui apresentado.

4.5.1. Camada de metadados

Conforme referido anteriormente, as iterações já concluídas não tiveram em conta a possibilidade de assentarem sobre uma camada de meta dados, transversal a todo o projeto BI-GEST. As capacidades de reporting providenciadas pelo atual projeto baseiam-se num acesso direto e incondicional ao DW, independentemente de ser por via do WAQR ou do Saiku Reporting.

No entanto, há várias vantagens em dotar este projeto de uma camada de meta dados, que funcionará como agente intermediário nas consultas e pedidos de dados que são realizados ao DW, permitindo oferecer ao utilizador final uma visão mais simples, amigável e flexível na construção dos seus próprios relatórios, conforme apresentado na seguinte figura:

Figura 49 - Apresentação das 3 camadas do Pentaho Metadata Editor (fonte: Pentaho)



Note-se, no entanto, que infelizmente a plataforma Pentaho não permite que as consultas OLAP sejam executadas em cima desta camada de meta dados, o que obrigará a um esforço adicional para replicar esta personalização diretamente na definição do schema Mondrian. Desta forma, apenas se retira vantagem da camada de meta dados para efeitos de reporting (tanto ao nível do Pentaho Report Designer, como para relatórios ad hoc no WAQR e Saiku Reporting).

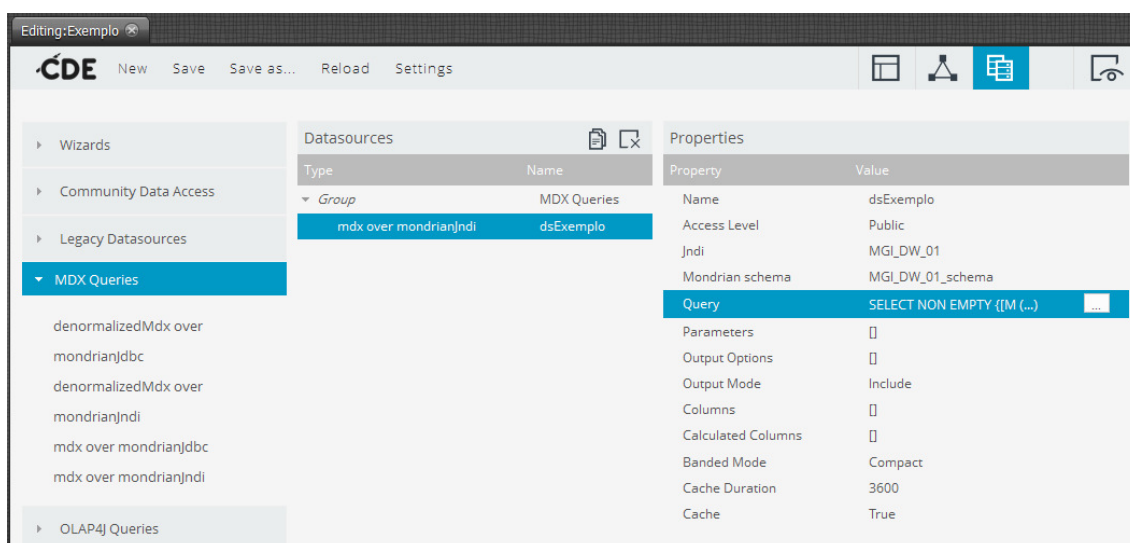
É importante salientar que o conceito de ‘camada de meta dados’, em termos de arquitetura Pentaho, se traduz na definição de 3 camadas distintas através da ferramenta utilitária Pentaho Metadata Editor (PME), a saber:

- **Camada física** – traduz-se na importação das estruturas físicas, onde ficam definidas as tabelas e colunas existentes no DW e que se constituirão como os blocos primários das camadas superiores;
- **Camada lógica** – permite redefinir as tabelas da camada anterior, bem como acrescentar colunas calculadas e estabelecer relações lógicas;
- **Camada de negócio** – é nesta camada que os objetos da camada lógica são agrupados e personalizados em unidades que fazem sentido aos utilizadores do sistema, permitindo que sejam acedidos e utilizados de forma fácil, amigável e de modo intuitivo;

4.5.2. Camada de Dashboards

Também faz parte dos planos para o BI-GEST o desenvolvimento de um conjunto relativamente abrangente de dashboards, desenvolvidos com recurso ao Community Dashboard Editor (CDE):

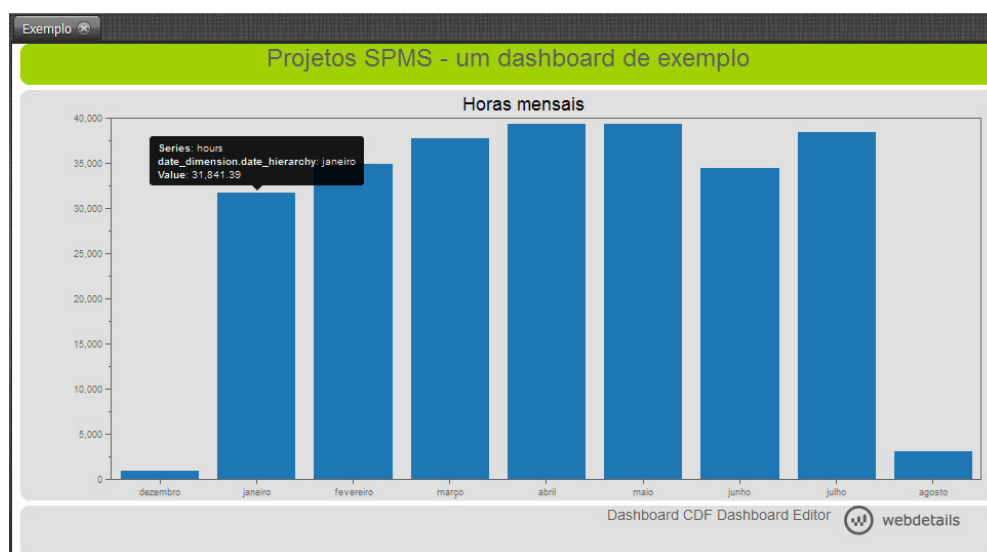
Figura 50 - Desenvolvimento de um dashboard no CDE (fonte: própria)



A disponibilização de dashboards aos utilizadores do BI-GEST possibilitará uma análise rápida dos diversos aspetos críticos inerentes ao progresso dos projetos informáticos na SPMS.

Note-se, no entanto, que embora a ferramenta CDE facilite significativamente a criação e edição de dashboards, permitindo que um pequeno gráfico de barras seja construído em questões de minutos, conforme demonstrado na figura seguinte:

Figura 51 - Desenvolvimento rápido de um dashboard de exemplo com CDE (fonte: própria)



Mesmo com o auxílio do CDE, o desenvolvimento de dashboards para utilização real exigem conhecimentos relativamente consolidados de vários domínios web: estruturação de conteúdos (html, html5), apresentação e estilos (css, blueprint), interatividade (javascript, xactions, ajax), entre outros.

4.5.3. Slowly Changing Dimensions

Um fator que não poderá ser descurado durante muito mais tempo no projeto BI-GEST prende-se com o facto de as dimensões estarem todas definidas de forma estática – ou seja, na escala utilizada originalmente por Kimball (Kimball & Ross, 2002), trata-se de uma SCD Tipo 1. No entanto, é sabido que os projetos de desenvolvimento de aplicações e soluções informáticas sofrem mutações constantes, com projetos a serem consolidados, separados, transformados, etc., ao mesmo tempo que a própria organização também revê a sua estrutura interna (novos departamentos, linhas d e serviços, etc.).

Por todas estas razões, cada uma das dimensões deste projeto necessitará de ser revista, para que se perceba qual a melhor estratégia a adotar em cada dimensão: SCD Tipo 1, Tipo 2 ou Tipo 3?

5. CONCLUSÕES

Após a execução de 4 ciclos completos de desenvolvimento, pensamos que a solução aqui documentada, para análise da gestão interna dos projetos informáticos da SPMS, esteja apta a ser avaliada. Relembramos que as principais motivações por detrás deste desenvolvimento – mais do que as funcionalidades de BI em si – eram a avaliação da viabilidade de plataformas open source, bem como ganhar experiência e poder analisar metodologias Ágeis de desenvolvimento de projetos de data warehouse e business intelligence.

Nesse sentido, importa agora referir algumas das limitações encontradas, os objetivos alcançados e, ainda, algumas possíveis recomendações práticas para os futuros projetos de business intelligence na SPMS.

5.1. LIMITAÇÕES

As principais limitações encontradas no âmbito deste projeto, especificamente no que concerne plataformas open source e metodologias Ágeis, foram as seguintes:

- i. **Grau de integração da plataforma.** Comparativamente com as ofertas comerciais que têm sido utilizadas pela SPMS (Oracle, Microsoft, Microstrategy, entre outros), a plataforma Pentaho não revela um grau tão forte de integração nos seus componentes. Sendo possível apresentar ao utilizador uma visão relativamente unificada da solução de BI, tal exige mais trabalho pela equipa técnica do que é o caso com produtos comerciais.
- ii. **Versatilidade e usabilidade da camada de exploração.** Também neste caso é notória a desvantagem da plataforma Pentaho, face às plataformas estabelecidas na SPMS. Mesmo considerando as melhorias significativas que resultam da instalação de componentes add-on, a capacidade de exploração avançada de dados é ainda um pouco limitada, por comparação com produtos comerciais.
- iii. **Grau de experiência na adoção de metodologias Ágeis.** Sendo inegavelmente uma forma mais flexível e dinâmica de desenvolvimento de projetos, sentimos que a adoção de metodologias Ágeis exige um maior grau de experiência no domínio do negócio e das próprias competências técnicas em conceitos de data warehousing e business intelligence. Caso contrário, parece haver riscos reais de dispersão de

esforços, e torna-se mais difícil coordenar a participação dos utilizadores em consonância com a visão do sponsor do projeto.

5.2. OBJETIVOS CONCRETIZADOS

Creemos que os dois principais objetivos deste trabalho foram alcançados:

- i. pensamos que tenha ficado provado que é, efetivamente, possível implementar projetos organizacionais de soluções de business intelligence com recurso exclusivamente a ferramentas open source. Ainda assim, pode acontecer que, no caso de requisitos muito específicos ou particularmente complexos, não haja plataforma open source compatível – algo que deve ser avaliado caso-a-caso.
- ii. também somos da opinião que as metodologias Ágeis, utilizadas no decorrer deste projeto, representam uma mais-valia para a gestão de projetos de business intelligence. Sendo certo que poderão exigir um maior grau de experiência profissional por parte da equipa técnica, a flexibilidade e dinamismo que acrescentam ao projeto representa uma mais-valia significativa.

5.3. RECOMENDAÇÕES FINAIS

Tendo em conta os objetivos atingidos, ainda que com as limitações já descritas, recomendamos que a SPMS considere adotar a plataforma de business intelligence Pentaho, na sua versão comunidade (sem custos de licenciamento), como alternativa viável para a generalidade dos seus projetos de BI, especialmente nos casos em que não apresentem requisitos específicos ao nível da camada de exploração (para esses casos, apenas um estudo individual permitirá otimizar a escolha de plataforma).

Simultaneamente, deve a SPMS considerar a incorporação de metodologias Ágeis no que concerne a gestão dos seus projetos informáticos, ainda que integrado na framework atualmente utilizada de PMBOK (que já reconhece e certifica este metodologia).

Neste sentido, será útil continuar a desenvolver algum trabalho exploratório para determinar quais os processos e práticas de trabalho, em concreto, que possam ou devam ser adotadas nos projetos da SPMS (e se em todos); tendo em conta a diversidade de metodologias específicas (Scrum, XP, etc.), será relevante levar a cabo mais alguma análise sobre a adequabilidade específica destas práticas, nas atividades diárias de gestão de projetos na SPMS.

6. BIBLIOGRAFIA

ACSS. (2010). *Plano Estratégico da ACSS*. Obtido de www.acss.min-saude.pt.

Ambler, S. W. (7 de Dezembro de 2011). *How Successful Are IT Projects, Really?* Obtido em 8 de Julho de 2012, de Dr. Dobb's: The World of Software Development: <http://www.drdoobbs.com/architecture-and-design/how-successful-are-it-projects-really/232300110#>

Bouman, R., & Dongen, v. J. (2009). *Pentaho Solutions: Business Intelligence and DataWarehousing with Pentaho and MySQL*. Indianapolis, EUA: Wiley Publishing, Inc.

Brooks, J. F. (1987). No Silver Bullet Essence and Accidents of Software Engineering. *Computer*, 20, 10-19.

Decreto-Lei n.º 108/2011, de 17 de novembro.

Decreto-Lei n.º 19/2010, de 22 de março.

Demarest, M. (junho de 1997). *The politics of data warehousing*. Obtido em novembro de 2014, de The University of North Carolina: <http://www.uncg.edu/ism/ism611/politics.pdf>

Devlin, B., & Murphy, P. (1988). An architecture for a business and information system. *IBM Systems Journal*, 60-80.

Evelson, B., & Hammond, J. (2010). *The Forrester Wave: Open Source Business Intelligence (BI), Q3 2010*. Forrester Research.

Faubert, J.-S. (junho de 2012). *A Review of Open Source Business Intelligence*. Obtido em novembro de 2013, de InEdge - Insurance Business Intelligence: <http://www.inedge.com/insurance-business-intelligence-and-analytics/documents/a-review-of-open-source-business-intelligence/>

Filipe, S. d., Bajanca, M., & Moreira, S. (2011). *Relatório Deloitte - Saúde em análise: Uma visão para o futuro*. Lisboa: Deloitte.

Fitzgerald, B. (1996). Formalized systems development methodologies: a critical perspective. *The Information Systems Journal*, Vol. 6, n.º 1, 3-23.

GNU. (s.d.). *A Quick Guide to GPLv3*. Obtido em novembro de 2014, de The GNU Operating System: <http://www.gnu.org/licenses/quick-guide-gplv3.html>

Gorman, W. (2009). *Pentaho Reporting 3.5 for Java Developers*. Birmingham, UK: Packt Publishing.

Hughes, R. (2013). *Agile data warehousing project management*. MA, USA: Elsevier.

Imhoff, C., Gallemmo, N., & Geiger, J. G. (2003). *Mastering Data Warehouse Design*. Indianapolis, EUA: Wiley Publishing, Inc.

Inmom, W. H. (2005). *Building the Data Warehouse*. Indianapolis: Wiley Publishing.

Kimball, R., & Ross, M. (2002). *The data warehouse toolkit : the complete guide to dimensional modeling*. New York: Wiley Computer Publishing.

Lindstrom, L. &. (2004). Extreme Programming and Agile Software Development Methodologies. *Information Systems Management* , 41-52.

Loshin, D. (2013). *Business intelligence: the savvy manager's guide*. MA, USA: Elsevier.

McNurlin, B., Sprague, R., & Bui, T. (2009). *Information systems management in practice*. Upper Saddle, NJ: Prentice Hall.

Negash, S. (2004, Volume 13). Business Intelligence. *Communications of the Association for Information Systems* , pp. 177-195.

Nerur, S., Mahapatra, R., & Mangalaraj, G. (2005). Challenges of Migrating to Agile Methodologies. *Communications of the ACM* , Vol. 48, n.º 5, 73-78.

PL 387/2013 - Proposta de Lei para o Orçamento de Estado para 2014.

Portal da Saúde. (6 de novembro de 2012). *Comunicação - Discursos e Intervenções*. Obtido em novembro de 2013, de Portal Da Saúde: <http://www.portaldasaude.pt/portal/conteudos/a+saude+em+portugal/ministerio/comunicacao/discursos+e+intervencoes/informatizacao+clinica+hospitalar+06112012.htm>

Pulvirenti, A. S., & Roldán, M. C. (2011). *Pentaho Data Integration 4 Cookbook*. Birmingham, UK: Packt Publishing.

Resolução do Conselho de Ministros n.º 12/2012.

Rodrigues da Silva, A. M., & Videira, C. A. (2005). *UML, Metodologias e Ferramentas Case Vol. I*. Lisboa: Centro Atlântico.

Saran, C. (s.d.). *Business intelligence software news*. Obtido em novembro de 2014, de Computer Weekly: <http://www.computerweekly.com/news/2240113585/Almost-a-third-of-BI-projects-fail-to-deliver-on-business-objectives>

SPMS. (julho de 2013). *Newsletter SPMS*. Obtido de www.spms.pt.

SPMS. (2012). *Orçamento da SPMS*. Obtido de www.spms.pt.

SPMS. (10 de julho de 2013). *Regulamento Interno da SPMS*. Obtido de www.spms.pt.

Vasiliev, Y. (2010). *Oracle Business Intelligence: The Condensed Guide to Analysis and Reporting*. Birmingham, UK: Packt Publishing.

Watson, H. J., & Wixom, B. H. (2007, Volume 9). The Current State of Business Intelligence. *COMPUTER - IEEE COMPUTER SOCIETY*, pp. 96-99.

Webdetails. (s.d.). *CTools - A story with truth in it*. Obtido em novembro de 2013, de infográfico da história da suite CTools: <http://www.webdetails.pt/info/storywithtruth.html>

Wieringa, R. (1998). A Survey of Structured and Object-Oriented Software Specification Methods and Techniques. *ACM Computing Surveys*, Vol. 30, n.º 4.

Wise, L. (2012). *Using Open Source Platforms for Business Intelligence*. MA, USA: Elsevier.